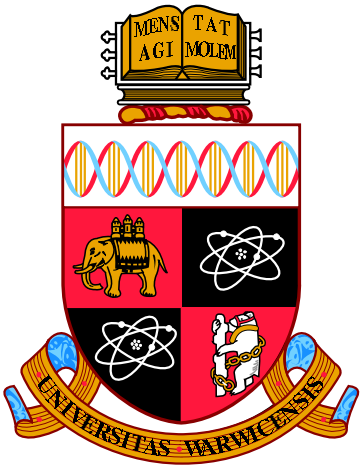

Three Essays in Econometrics

by

Andrea Anita Naghi

Thesis

Submitted to the University of Warwick

for the degree of

Doctor of Philosophy

Department of Economics

September 2016

THE UNIVERSITY OF WARWICK

Contents

List of Tables

List of Figures

Acknowledgments

I would like to thank my supervisors: Valentina Corradi, Ana Galvao, Ivana Komunjer and Jeremy Smith for their continuous encouragement and support throughout my PhD studies. I am extremely grateful for all their guidance, invaluable suggestions and comments on various previous drafts, and for always being so helpful. Financial support from the Economic and Social Research Council as well as the Royal Economic Society is thankfully acknowledged. Part of this thesis was written while the author was visiting the UCSD Economics Department, whose warm hospitality is gratefully acknowledged. Special thanks to my friends and colleagues, especially: Tina Ammon, Anna Baiardi, Mahnaz Nazneen, Andis Sofianos, Spyros Terovitis and Raghul Venkatesh who were a source of constructive discussions and continuous solid support. Finally, I would like to thank my parents: Elza and Jozsef and my brother Attila for always believing in me.

Declarations

I submit this thesis to the University of Warwick in accordance with the requirements for the degree of Doctor of Philosophy. I declare that this thesis is my own work and has not been submitted for a degree at another university.

Andrea Anita Naghi

September 2016

Abstract

Chapter One, *A Forecast Rationality Test that Allows for Loss Function Asymmetries*, proposes a new forecast rationality test that allows for asymmetric preferences without assuming any particular functional form for the forecaster's loss function. The construction of the test is based on the simple idea that under rationality, asymmetric preferences imply that the conditional bias of the forecast error is zero. The null hypothesis of forecast rationality under asymmetric loss (i.e. no conditional bias) is tested by constructing a Bierens [1982, 1990] conditional moment type test.

Chapter Two, *Global Identification Failure in DSGE Models and its Impact on Forecasting*, considers the identification problem in DSGE models and its transfer to other objects of interest such as point forecasts. The results document that when observationally equivalent parameter points belong to the same model structure, the implied point forecasts are the same under both correct specification and misspecification. However, when analyzing the identification problem permitting models with different structures (e.g. different policy rules that produce sets of data dynamics that are quantitatively similar), the paper shows that indistinguishable parameter estimates can lead to distinct predictions.

Chapter Three, *Identification Robust Predictive Ability Testing*, considers the predictive accuracy evaluation of models that are strongly identified in some part of the parameter space but non-identified or weakly identified in another part of the parameter space. The paper shows that when comparing the predictive ability of models that might be affected by identification deficiencies, the null distribution of out-of-sample predictive ability tests is not well approximated by the standard normal distribution. As a result, employing a standard (strong) identification critical value might lead to misleading inference. We propose methods to make the out-of-sample predictive ability tests robust to identification loss.

Chapter 1

A Forecast Rationality Test that Allows for Loss Function Asymmetries

1.1 Introduction

The choice of loss function is an important problem in the forecast assessment literature as it has a direct effect on the results of forecast optimality tests. Traditionally, empirical studies have based their forecast evaluation framework on the assumption of a symmetric loss, which implies that positive and negative forecast errors are equally weighted by the forecaster. Among the symmetric loss functions, the one that is the most frequently used is the mean squared error loss, under which forecast efficiency has been studied by testing whether the forecast errors have zero mean or whether they are uncorrelated with the available information at the time the forecast was made. A common practice for testing forecast unbiasedness, for example, is to rely on early works such as

Theil [1958], Mincer and Zarnowitz [1969] and construct a regression of realized values on a constant and the forecast, testing whether the implied coefficients are zero and one respectively. Nevertheless, the assumption of a symmetric loss can be difficult to justify sometimes and often it is not plausible. For example, at a macroeconomic level, central banks can be averse to bad outcomes such as lower than expected real output growth and higher than expected inflation and hence they incorporate this loss aversion into their forecasts. At the level of individual firms, the cost of under-predicting demand, which results in loss of sales, should not necessarily be the same as the cost of over-predicting demand, which means additional storage costs.

Given that symmetric loss functions such as the mean squared error or mean absolute error may not be flexible enough to capture the loss structures that forecasters face, another line of the literature (see e.g. Christoffersen and Diebold [1997], Elliott et al. [2005, 2008] (EKT hereafter), Patton and Timmermann [2007a], Patton and Timmermann [2007b], Komunjer and Owyang [2012]), argues that an asymmetric loss function that weights differently positive and negative forecast errors could be more representative for the forecaster's intentions. However, under an asymmetric loss, standard properties of optimal forecast are no longer valid and the traditional forecast rationality tests could be misleading, not being able to distinguish whether the forecasters use inefficiently their information, or whether the underlying loss function is just asymmetric. Thus, rejections of rationality in the standard rationality evaluation literature may largely be caused by the assumption of a squared loss function. Consequently, this strand of the literature points to the need to develop testing procedures that are robust to a broader class of loss functions.

Important progress has been made in this direction. Patton and Tim-

mermann [2007a], established properties of optimal forecasts valid under the Linex loss function [1] and a popular nonlinear DGP - the regime switching model of Hamilton [1989]. Patton and Timmermann [2007b] propose tests for forecast optimality that do not require the knowledge of the specific loss function, but some testable restrictions have to be imposed on the DGP. Another framework that is flexible regarding the forecaster's loss function is the one in Elliott et al. [2005, 2008]. They provide a GMM-based forecast optimality testing framework based on a general class of loss functions that allows for a parametrization of the asymmetry in the loss and includes the quadratic loss as a special case. As a by-product of the test, an estimate of the asymmetry parameter of the loss function is obtained. Komunjer and Owyang [2012] extend the framework of EKT to a new family of multivariate loss, which by construction does not impose independence across variables in the loss function. The framework allows for asymmetries in the forecaster's loss and permits testing the rationality of a vector of forecasts. More recently, the importance of evaluating forecasts under the loss function that is consistent for the functional of interest (mean, quantile, distribution) has been brought into attention (see Gneiting [2011], Patton [2014]). For example, if the functional of interest that is forecasted is the mean, the forecasts should be evaluated under the general class of Bregman loss functions (Bregman [1967]). Interestingly, under an asymmetric loss function that belongs to the Bregman class, the optimal forecast for the mean should not necessarily be biased (see Patton [2014]).

This paper contributes to the literature of forecast assessment under asymmetric loss. We establish a testable property for forecast rationality that

[1] The Linex loss function is chosen because it is a popular way to represent asymmetric preferences.

holds when the forecaster's loss function is unknown but it is assumed to be asymmetric, the framework being also able to accommodate the particular case when the loss is unknown but symmetric. Our paper is most closely related to Elliott et al. [2005, 2008]. Unlike Elliott et al. [2005, 2008], our approach accounts for the possibility of asymmetry, without restricting the forecaster's loss to any particular parametric form. The attractiveness of our approach from a practical point of view, is that it can be applied even if the forecast user does not have any information regarding the shape of the forecaster's loss function. Furthermore, we should note that in the construction of our test statistic, we require neither the knowledge of the underlying loss function nor the knowledge of the forecasting model used by the forecaster. The forecast errors required to compute the test could have been generated by a parametric, nonparametric, semiparametric, or no model at all. The construction of the proposed test is drawn on the conditional moment tests of Bierens [1982, 1990], De Jong [1996], Corradi and Swanson [2002], Corradi et al. [2009].

Section 1.2 provides a brief review of a commonly used framework for testing forecast rationality under asymmetry and points out some of its limitations. In Section 1.3 we establish a testable property for forecast rationality under asymmetric loss and outline our proposed test. In Section 1.4 we present the advantages of our framework by studying the finite sample properties of our test under a nonlinear DGP, and emphasizing the effect of a misspecified loss in the framework of EKT. Section 1.5 presents an empirical illustration using data from the Survey of Professional Forecasters (SPF). Concluding remarks are provided in Section 2.5. The main simulation and empirical findings are presented in Appendix A1. Appendix A2 contains additional empirical results that support our findings.

1.2 Testing Forecast Rationality under Asymmetric Loss: the EKT Framework

Elliott et al. [2005, 2008], address the issue of allowing for asymmetric preferences when testing forecast optimality. The forecast rationality test that they propose is based on a pre-specified class of loss functions that has the following parametric functional form:

$$L_1\ (\epsilon_{t+1}; p, \alpha) = [\alpha + (1 - 2\alpha) \cdot 1(\epsilon_{t+1} < 0)] \cdot |\epsilon_{t+1}|^p \tag{1.1}$$

This function depends on the forecast error ϵ_{t+1} and the shape parameters $p \in \mathbb{N}^*$ and $\alpha \in (0, 1)$.

The parameter α describes the degree of asymmetry in the forecaster's loss function. This parameter is of important economic interest as it provides information about the forecaster's objectives that can be useful for forecast users. For values of α less than one half the forecaster gives higher weights on negative forecast errors than on positive ones of the same magnitude, or in other words over-prediction is more costly than under-prediction. Values greater than one half indicate a higher cost associated with positive forecast errors, or that under-prediction is more costly than over-prediction. In the symmetric case, α equals one half, in which case the costs associated with positive and negative forecast errors are equally weighted. The relative cost of a forecast error can be estimated as $\hat{\alpha}/1 - \hat{\alpha}$. For example, if the estimated value of α is 0.75, positive forecast errors obtained by under-forecasting are three times more costly then negative ones obtained by over-forecasting (see e.g. Capistrán [2008]).

Special cases of L_1 include the absolute deviation loss function $L_1(\epsilon_{t+1}; 1, 1/2)$ the squared error loss function $L_1(\epsilon_{t+1}; 2, 1/2)$ and their asymmetrical counterparts obtained when $\alpha \neq 1/2$ - the lin-lin loss $L_1(\epsilon_{t+1}; 1, \alpha)$ and the quad-quad loss $L_1(\epsilon_{t+1}; 2, \alpha)$. Thus, while this class of loss functions allows for asymmetric preferences, it also nests the popular symmetric loss functions widely used in empirical studies.

A sequence of forecasts is said to be optimal under a particular loss function if the forecast minimizes the expected value of the loss, conditional on the information set of the forecaster. Elliott et al. [2005, 2008], exploit the idea that optimal forecasts must satisfy the first order condition of the forecaster's optimization problem, and construct a forecast rationality test based on the following moment conditions:

$$E(W_t(1(\epsilon^*_{t+1} \leq 0) - \alpha_0)|\epsilon^*_{t+1}|^{p_0 - 1}| = 0 \tag{1.2}$$

where $\epsilon^*_{t+1} = y_{t+1} - f^*_{t+1}$ is the optimal forecast error, the difference between the realization of some target variable Y_{t+1} and the optimal forecast, p_0 and α_0, are the unknown true values of p and α, and W_t is the information set of the forecaster.

Using the moment conditions in (1.2), first a GMM estimator for the asymmetry parameter α is obtained. Then, whether the moment conditions in (1.2) associated with the first order condition of the forecaster's optimization hold (i.e. the null hypothesis of forecast rationality), is tested by constructing

a test for overidentification:

$$J = \frac{1}{T}[\sum_{t=\tau}^{T+\tau-1} v_t[1(\hat{e}_{t+1} < 0) - \hat{\alpha}_T]|\hat{e}_{t+1}|^{p_0-1}]'\hat{S}^{-1} \\ [\sum_{t=\tau}^{T+\tau-1} v_t[1(\hat{e}_{t+1} < 0) - \hat{\alpha}_T]|\hat{e}_{t+1}|^{p_0-1}] \tag{1.3}$$

The test is asymptotically distributed as a χ^2 with $d-1$ degrees of freedom, with d the size of the vector of instruments V_t, and rejects for large values. In (1.3), $\hat{e}_{t+1}$ is the observed forecast error, τ is the estimation sample size (in-sample size), T is the number of forecasts available (out-of-sample size), v_t are the observations of the vector of instruments V_t, $\hat{\alpha}_T$ is a linear instrumental variable estimator of the true value α_0,

$$\hat{\alpha}_T \equiv \frac{[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t|\hat{e}_{t+1}|^{p_0-1}]'\hat{S}^{-1}[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t 1(\hat{e}_{t+1} < 0)|\hat{e}_{t+1}|^{p_0-1}]}{[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t|\hat{e}_{t+1}|^{p_0-1}]'\hat{S}^{-1}[\frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t|\hat{e}_{t+1}|^{p_0-1}]} \tag{1.4}$$

and $\hat{S}$, defined as $\hat{S}(\bar{\alpha}_T) \equiv \frac{1}{T}\sum_{t=\tau}^{T+\tau-1} v_t v_t'(1(\hat{e}_{t+1} < 0) - \bar{\alpha}_T)^2|\hat{e}_{t+1}|^{2p_0-2}$, is a consistent estimate of a positive definite weighting matrix S, depending on $\bar{\alpha}_T$, a consistent initial estimate for α_0.

While this approach takes into account asymmetric preferences and allows for a great flexibility regarding the loss function, it relies on a couple of assumptions. First, it maintains the assumption that the loss function belongs to the parametric form given in (1.1). This particular functional form is substituted in the forecaster's minimization problem and then the moment conditions in (1.2) are obtained accordingly. The problem is that if the forecaster's true loss function does not belong to this parametrization, the test could lead to misleading inferences. Our analysis will show this in section 1.4.

Second, the method is based on the assumption that the forecasts were generated using a linear model of the type: $f_{t+1} = \theta' W_t$, where θ is a k-vector of parameters in a compact set $\Theta \subset R^k$, and thus, the observed forecast error is: $\hat{e}_{t+1} = y_{t+1} - \hat{\theta}' W_t$. In this framework, failure to reject the null hypothesis of rationality, means an absence of linear correlation between the information set of the forecaster and the forecast error. Hence, possible nonlinear dependencies are not necessarily detected. The forecast error could be uncorrelated with W_t but correlated with a nonlinear function of W_t. Moreover, the error could be correlated with some variables not even included in W_t. In order to apply this approach in the case of a nonlinear forecasting rule, $f_{t+1} = f(\theta, W_t)$, W_t in (1.2) would have to be replaced with the gradient of f with respect to the parameter θ evaluated at (θ^*, W_t). However, the forecasting model f, its true parameters θ^* and W_t are usually not known by the forecast user.

1.3 Forecast Rationality Tests under Unknown Functional Form of the Loss

In this section, we propose a testable property for forecast rationality under asymmetric loss and outline a forecast rationality test that can be used when the forecast evaluator wishes to test for forecast optimality suspecting that the forecasters give different weights to positive and negative forecast errors. The proposed test is not based on the assumption of a particular parametric form for the loss function, the asymmetric preferences being captured in a non-parametric way. This constitutes an important feature of the test, as the true loss function used in the construction of the forecasts is not known in forecast

evaluation. Another attribute of the test is that it detects possible nonlinear dependencies between the forecast error and the forecaster's information set, therefore it allows for a nonlinear forecasting rule.

To this end, consider first the following null hypothesis

$$H_0 : E(\epsilon_{t+1}|W_t) = 0 \tag{1.5}$$

against the alternative

$$H_1 : E(\epsilon_{t+1}|W_t) \neq 0$$

where ϵ_{t+1} is the forecast error and W_t contains all publicly available information relevant to predict a variable Y_{t+1} at time t. If the null in 1.5 is true, it means *rationality* and *unbiasedness* of the forecast error, as the forecast error follows a martingale difference sequence (m.d.s.) with respect to the information set of the forecaster.

Define now

$$H_0 : E(\epsilon_{t+1}|W_t) = E(\epsilon_{t+1}) \tag{1.6}$$

against

$$H_1 : E(\epsilon_{t+1}|W_t) \neq E(\epsilon_{t+1})$$

The null in (1.6) implies that the conditional expectation of the forecast error equals the unconditional expectation and thus the forecast error is independent of any function which is measurable in terms of the information set available at time t. In other words, there is no issue of inefficient use of the available information and thus the forecasts are *rational*. Now, if (1.6)

is true and $E(\epsilon_{t+1}|W_t) \neq 0$ the forecast errors are not only rational but also *biased*. Thus, (1.6), under the assumption that $E(\epsilon_{t+1}|W_t) \neq 0$, constitutes a null hypothesis for testing forecast rationality under asymmetric loss.

The null given in (1.6) can be now restated as follows: $H_0 : E(\epsilon_{t+1}|W_t) - E(\epsilon_{t+1}) = 0 \Leftrightarrow H_0 : E(\epsilon_{t+1}|W_t) - E[E(\epsilon_{t+1})|W_t] = 0 \Leftrightarrow H_0 : E[\epsilon_{t+1}|W_t - E(\epsilon_{t+1}|W_t)] = 0 \Leftrightarrow$

$$H_0 : E[(\epsilon_{t+1} - E(\epsilon_{t+1}))|W_t] = 0 \tag{1.7}$$

The alternative of the new form of the null hypothesis can be written as:

$$H_1 : Pr\ [E[(\epsilon_{t+1} - E(\epsilon_{t+1}))|W_t] = 0] < 1$$

From the new form of the null hypothesis given in (1.7), it can be seen that testing for forecast rationality under asymmetric loss, reduces to testing whether the quantity $E[(\epsilon_{t+1} - E(\epsilon_{t+1}))|W_t]$, i.e. the *conditional bias of the forecast error*, is zero.

We are now able to construct a Bierens [1982, 1990] type test for the null given in (1.7).

To this end, we apply to our context the test statistic suggested by De Jong [1996] which generalizes the consistent model specification test proposed by Bierens [1990], to the case of data dependence. Thus, we define:

$$M_T = sup_{\gamma \in \Gamma} |m_T(\gamma)| \tag{1.8}$$

where:

$$m_T(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} (\hat{e}_{t+1} - \overline{e})\, w \left(\sum_{j=0}^{t-1} \gamma_j' \Phi(W_{t-j}) \right) \tag{1.9}$$

Exploiting the equivalence $E[(\epsilon_{t+1} - E(\epsilon_{t+1}))|W_t] = 0 \Leftrightarrow E[(\epsilon_{t+1} - E(\epsilon_{t+1}))w(\gamma', W_t)] = 0$, consistent model specification tests are based on the discrepancy of the sample analog of $E[(\epsilon_{t+1} - E(\epsilon_{t+1}))w(\gamma', W_t)]$ to zero. In (1.9), $\hat{e}_{t+1}$ is the observed one-step ahead forecast error obtained as the difference between the actual realization and the forecasted value from the forecast producer, $\hat{e}_{t+1} = y_{t+1} - \hat{f}_{t+1}$. The mean $\overline{e}$ is defined as $\overline{e} = \frac{1}{T} \sum_{t=0}^{T-1} \hat{e}_{t+1}$, and T is the number of observed forecast errors. Our discussion focuses on $\hat{e}_{t+1}$, but the results generalize to $\hat{e}_{t+h}$, where $h > 1$ is the forecast horizon.

The function $w(\gamma', W_t)$ is a generically comprehensive function, a nonlinear transformation of the conditioning variables. Commonly used functions in the literature for w are: $w(\gamma', W_t) = \exp(\sum_{i=1}^{k} \gamma_i \Phi(W_{i,t}))$, or $w(\gamma', W_t) = 1/(1 + \exp(c - \sum_{i=1}^{k} \gamma_i \Phi(W_{i,t})))$, where c is a constant, $c \neq 0$, and Φ is a measurable one-to-one mapping from R to a bounded subset of R, it can be chosen the arctangent function, for example. The choice of the exponential in the weight function w is not crucial. Stinchcombe and White [1998] show that any function that admits an infinite series approximation on compact sets with non-zero series coefficients can be used to obtain a consistent test. The weights, γ_j, attached to observations decrease over time.

The following remarks are worth making at this point.

Remark 1. In the particular case in which forecast efficiency is tested under a symmetric loss function, one can test $H_0 : E(\epsilon_{t+1}|W_t) = 0$, which if it is true it means that the forecast error is unbiased and follows a m.d.s. with respect

to the information set used by the forecaster. The test statistic thus becomes: $M_T = sup_{\gamma\in\Gamma}|m_T(\gamma)|$, with $m_T(\gamma) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}\hat{e}_{t+1}\, w\left(\sum_{j=0}^{t-1}\gamma_j'\Phi(W_{t-j})\right)$.

Remark 2. Suppose now that the true loss function is asymmetric and one uses $M_T = sup_{\gamma\in\Gamma}|m_T(\gamma)|$, with $m_T(\gamma) = \frac{1}{\sqrt{T}}\sum_{t=0}^{T-1}\hat{e}_{t+1}\, w\left(\sum_{j=0}^{t-1}\gamma_j'\Phi(W_{t-j})\right)$ to test for forecast rationality, then this form of the test might falsely reject rationality as it is based on the assumption of a symmetric loss.

The proposed test statistic has a limiting distribution that is a functional of a Gaussian process (see e.g. Corradi et al. [2009]). Under H_0, $M_T \xrightarrow{d} \sup_{\gamma\in\Gamma}|m_T(\gamma)|$, where $m(\gamma)$ is a zero mean Gaussian process. Under H_1, there exist an $\epsilon > 0$, such that: $Pr\left(\frac{1}{\sqrt{T}}M_T > \epsilon\right) \to 1$. The proof follows from the empirical process CLT of Andrews [1991], for heterogeneous near-epoch dependent (i.e. functions of mixing processes) arrays. The limiting distribution of the statistic, M_T is the supremum over a Gaussian process and hence standard critical values are not available. Also, note that M_T is not pivotal because the limiting distribution depends on the nuisance parameter $\gamma \in \Gamma$. The test has power against generic nonlinear alternatives, but the critical values have to be computed by bootstrap. In the Monte Carlo study and the empirical part of the paper, the block bootstrap is employed to obtain the critical values for the test. In the block bootstrap, $\hat{e}_{t+1}$ and the data are jointly resampled in order to preserve the correct temporal behavior and to mimic the original statistic.

1.4 Monte Carlo Evidence

Our Monte Carlo study consists of two parts. First, we emphasise the importance of the choice of the loss function in forecast evaluation. To this end, we

illustrate the effect that a misspecified loss function can have in the forecast evaluation framework of EKT. Then, we compare the empirical power of our proposed test with that of the J-test, in the presence of nonlinear dependencies between the sequence of forecast errors and the information set available at the time the forecast is made.

1.4.1 The Effect of a Misspecified Loss Function in Forecast Evaluation

We illustrate the loss function sensitivity of the EKT framework, by constructing a Monte Carlo exercise where the forecaster's true loss function does not belong to the class of loss functions on which the test is based on. Nevertheless, the forecast evaluation is done under the particular class of loss introduced in EKT. To highlight the effect of a misspecified loss function, we examine the behavior of the estimator $\hat{\alpha}$ and study the properties of the J-test.

We assume that the variable of interest is generated by a simple AR(1) process:

$$X_t = b + cX_{t-1} + \epsilon_t$$

where the errors are serially uncorrelated, $\epsilon_t \sim N\ (0,\ 0.5)$, and the parameters are set to $b = 0.9$ and $c = 0.7$. We generate random samples of size $T = R + P - 1$, after discarding the first 100 observations to remove initial values effects. Using a rolling window of size R, the forecaster constructs P one period ahead forecasts by minimizing the expected value of the loss function, L_1, assumed to be of the form given in (1.1). The observed one period ahead

forecast is thus $\hat{f}_{t+1,t} = \hat{b} + \hat{c}X_t$ where $\hat{b}$ and $\hat{c}$ are obtained by minimizing L_1:

$$(\hat{b}, \hat{c}) = \arg\min\ R^{-1} \sum_{t=1}^{R} L_1\ (\alpha_0,\ X_{t+1} - b - cX_t) \tag{1.10}$$

where α_0 is the true value of the forecaster's loss function asymmetry parameter. The sequence of the observed forecast errors is then computed as:

$$\{\hat{e}_{t+1}\}_{t=R}^{T} = \{X_{t+1} - \hat{b} - \hat{c}X_t\}_{t=R}^{T}$$

We perform 1000 Monte Carlo simulations for different choices of R, P and α_0. The instrument set in this section includes a constant and the lagged forecast errors.

Table A.1 reports the average α_0 estimates for various sample sizes and various values of the true asymmetry parameter. As expected, the estimator performs overall well when the loss function is correctly specified, the estimated values being close to the true values. Table A.2 reports the empirical rejections probabilities for the J-test. Under a correctly specified loss, size is well controlled. There are some size distortions in cases when the in-sample size is smaller or equal to the out-of-sample size, indicating the importance of controlling the relative sizes of R and P.

Now, we examine the implications of falsely assuming that the forecaster's true loss function belongs to (1.1). We reconstruct our Monte Carlo exercise assuming now that the forecaster's true loss function is the Linex loss (see Varian [1975], Zellner [1986]), an asymmetric loss that involves both linear

and exponential terms and it is defined as:

$$L_2(\epsilon_{t+1}; a) = \exp(a \cdot \epsilon_{t+1}) - a \cdot \epsilon_{t+1} - 1 \tag{1.11}$$

This time, the forecast evaluation is done inaccurately under the loss function given in (1.1). In this case, $\hat{b}$ and $\hat{c}$ are obtained as:

$$(\hat{b}, \hat{c}) = \arg\min \; R^{-1} \sum_{t=1}^{R} L_2 \; (a_0, \; X_{t+1} - b - cX_t)$$

where a_0 is the true value of the Linex loss function's asymmetry parameter.

Table A.3 reports the average GMM estimates of α for different sample sizes and different values of the Linex loss function's true asymmetry parameter. The average estimates present large variations across different values of the true loss function's asymmetry parameter, and they are far from the true values. The results given in Table A.4 indicate the size distortions of the J-test obtained when the forecast evaluation is done under a misspecified loss function. The J-test tends to over reject the null of rationality, the size distortions being larger for larger values (in absolute value) of the Linex loss function's asymmetry parameter. The size distortions seems to be smaller though in cases where the Linex loss function's parameter is closer to the origin.

In order to better understand when the J-test suffers from size distortions, figures A.1 to A.6, plot the EKT and the Linex loss functions on the same graph with different parametrizations. Take for example Figure A.1. Here, the Linex loss function's parameter, a, is fixed to 1 while the EKT loss function's parameters are $p = 2$ and $\alpha = 0.65$ - the incorrectly estimated value for α obtained in Table A.3, when the Linex loss true parameter, a_0, was 1.

Under these parametrizations, the two loss functions overlap on a notably large region and the rejection frequencies for the J-test, under a misspecified loss function, (as obtained in Table A.4), are not that far from the nominal level of 5%. Figure A.3, shows a case when the two functions have only a very small overlapping region at the origin. In this graph, $a = 3$, $p = 2$ and $\alpha = 0.86$. When the common region of the two functions is small, the J-test does not control size well and leads to over rejections of the null of rationality.

1.4.2 Empirical Size and Power Comparison

In the second Monte Carlo setup, we consider the following data generating process:

$$Y_t = \theta' W_t + \delta g(\phi' W_t) + U_t \tag{1.12}$$

where $\theta' = (\theta_1, \theta_2)$, $\phi' = (\phi_1, \phi_2)$ and $W_t = (W_{1t}, W_{2t})'$. We set the following expressions for the nonlinear function g, $g(x) = x^2$, $g(x) = exp(x)$, $g(x) = arctan(x)$. Different parametrization are set for δ, the parameter that quantifies the size of the nonlinearity, as indicated in Table A.5. The parameters governing the process, θ and ϕ, are fixed to $(\theta_1, \theta_2) = (0.5, 0.5)$, $(\phi_1, \phi_2) = (0.7, 0.8)$. In this section, the instrument set is a constant and Z_t - which is generated such that it is correlated with W_{1t}, W_{2t} but uncorrelated with U_t. In order to ensure this, we generate W_{1t}, W_{2t}, Z_t, U_t from a multivari-

ate normal distribution:

$$\begin{pmatrix} W_{1t} \\ W_{2t} \\ Z_t \\ U_t \end{pmatrix} \sim N(0, \Sigma)$$

where Σ is the variance-covariance matrix set to:

$$\Sigma = \begin{pmatrix} 2 & 0.1 & 0.8 & 0.2 \\ 0.1 & 2 & 1 & 0.1 \\ 0.8 & 1 & 2 & 0 \\ 0.2 & 0.1 & 0 & 0.8 \end{pmatrix}$$

We generate a sample of size $T = 500$ for Y_t according to the data generating process given in 1.12. We assume the forecaster uses the first $R = 0.6 \times T$ observations to estimate the parameters of a linear model:

$$Y_t = \theta_1 W_{1t} + \theta_2 W_{2t} + U_t$$

The sequence of one-step ahead forecasts, $\hat{Y}_{t+1} = \hat{\theta}_1 W_{1t} + \hat{\theta}_2 W_{2t}$, and the observed one-step ahead forecast errors $\hat{e}_{t+1}$ are obtained using a recursive scheme. Given that a linear forecasting model was used to generate the forecast errors even though the data was generated by a nonlinear process, we ensure that the resulting forecast error is correlated with some nonlinear function of W_t. This means that whenever $\delta \neq 0$, we can study the empirical power of the tests, while when we set δ to 0, we obtain the empirical sizes. We perform 1000 Monte Carlo simulations for both statistics. In addition, for each Monte

Carlo simulation we perform 100 bootstrap simulations in order to obtain the critical values for the M_T statistic.

Table A.5 reports the rejection frequencies at a 10% significance level for different nonlinear functions, $g(x)$, and for different parametrization for δ. When we set $\delta = 0$, the forecast error is uncorrelated with the available information and both tests have an empirical size close to the nominal size of 10%. In all the other cases characterized by a nonlinear relationship between the forecast error and the information set, the M_T test outperforms the J-test.

1.5 Empirical Illustration

In this section, we perform an empirical comparison of the M_T and J tests using data from the Survey of Professional Forecasters (SPF) maintained by the Federal Reserve Bank of Philadelphia. In this data set, survey participants provide point forecasts for macroeconomic variables in quarterly surveys. As the objective of the forecaster is unknown, it is not sure that the forecaster simply minimizes a quadratic loss function and reports the conditional mean. Thus, when evaluating these forecasts, it is reasonable not to impose too much structure on the unknown loss function. Nevertheless, the reported forecasts should indeed reflect the underlying loss function.

The following series were selected for this empirical application: quarterly growth rates for real GNP/GDP (1968:4-2012:4),[2] the price index for GNP/GDP (1968:4-2012:4), and the quarterly growth rates for consumption (1981:3-2012:4).[3] The growth rates are calculated as the difference in natural logs. For robustness, we perform our analysis on median, mean and range

[2]The SPF provides data on GNP before 1992 and on GDP after 1992.
[3]Labeled as Real Personal Consumption Expenditures in the SPF data set.

forecasts. Results on the median mean and range forecasts are provided in Appendix A.

In the computation of the two statistics, we considered the one-step ahead forecast error obtained as the difference between the actual and the one-step ahead forecasted value. The point forecast data set of the SPF provides data on the year, the quarter, the most recent value known to the forecasters, the value for the current quarter (which is usually forecasted) and then forecasts for the next four quarters. To compute the one-step ahead forecasted growth rates, we used values corresponding to the current quarter and the most recent known value. For the actual values, the SPF provides a real-time data set. In order to compute the actual growth rates we used the first release.

For the instruments of the J-test and the information set used in the computation of the M_T test, we considered the following six cases: Case 1 - constant and lagged errors, Case 2 - constant and absolute lagged errors, Case 3 - constant and lagged change in actual values, Case 4 - constant and lagged change in forecasts, Case 5 - constant, lagged errors and lagged change in actual values, Case 6 - constant, lagged errors, lagged change in actual values and lagged change in forecasts.

In the construction of the M_T statistic we adopt commonly used settings in the literature. We chose the exponential function for w and the arctangent function for Φ. Also, we set $\gamma_j \equiv \gamma(j+1)^{-2}$, where $\gamma \in [0,3]$ when γ is unidimensional - as it is for Case 1. When γ is multidimensional, we have for

example for Case 5:

$$\gamma = \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \end{pmatrix} \in [0,3] \times [0,3] \times [0,3]$$

and the test statistic is computed as the supremum of the absolute value of:

$$\begin{aligned} m_T(\gamma) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} (\hat{e}_{t+1} - \overline{e}) \exp \Bigg[\sum_{j=0}^{t-1} (\gamma_1 (j+1)^{-2} \tan^{-1}(Z_{1,t-j}) \\ + \gamma_2 (j+1)^{-2} \tan^{-1}(Z_{2,t-j}) + \gamma_3 (j+1)^{-2} \tan^{-1}(Z_{3,t-j})) \Bigg] \end{aligned} \tag{1.13}$$

where Z_1 is a vector of ones, Z_2 contains the lagged errors and Z_3 the lagged change in actual values. The critical values for the M_T test, are computed using the block bootstrap with blocks of length 5 and an overlap length of 2. Given the small sample sizes, we derive our conclusions based on the 10% bootstrap critical values.

Table A.6 reports the results obtained for the J-test based on the median forecasts. The estimates of the asymmetry parameter for the real GNP/GDP take values slightly less than 0.5, while for the price index the estimates take values slightly higher than 0.5. However, when performing a t-test that tests $H_0 : \alpha = 0.5$, the null of symmetric preferences, in the context of the EKT class of loss, cannot be rejected for GNP/GDP and the price index. Interestingly, this null hypothesis is rejected for consumption - variable for which forecasters tend to under-predict. At the 10% level, the J-test does not reject the composite null hypothesis that the loss belongs to the family of loss functions defined in EKT and that the forecasts are rational for GDP/GNP

and the price index. However, it does reject for most instrument set cases of consumption - specifically for Cases 1, 3, 5 and 6.

Analyzing now Table A.7, where our suggested nonlinear test is computed, we notice that for the real GNP/GDP, our test results are in conformity with the J- test's results - forecast rationality is not rejected for this variable. We obtain however contrasting results that reveal interesting insights for the price index and consumption. Unlike the J-test, the M_T test rejects forecast rationality for the price index, which suggests that in the case of inflation, the forecast error depends in a nonlinear fashion on the information set used to produce the forecasts and the J-test is not able to detect these nonlinear dependencies. For consumption, our test does not reject rationality, even though the J-test rejects the null. This might be an indication that the true loss function used to generate the forecasts for consumption was from a different family of loss than the one the J-test is based on, and consequently the J-test cannot distinguish whether the forecasters use inefficiently their information or whether the underlying loss function does not belong to the pre-specified loss, and therefore it rejects the null.

The empirical results on the mean and the range responses confirm our results on median forecasts. The results on these two series are included in Appendix A.

1.6 Conclusion

In this paper, we propose a new forecast rationality test that allows for asymmetric preferences, without assuming a particular functional form for the forecaster's loss function. The key idea in the construction of our test is that fore-

cast rationality under an asymmetric loss function implies a zero conditional bias of the forecast error. Our framework is based on the classical literature on consistent model specification tests in the spirit of Bierens [1982, 1990], De Jong [1996]. The asymmetry in the loss function is captured in a nonparametric way. The drawback of the latter is that in contrast to Elliott et al. [2005, 2008], this approach cannot quantify the magnitude of the asymmetry in the loss function, which can be of important economic interest.

Monte Carlo simulations illustrate the advantages of our approach. We show that a commonly used test that accounts for the possibility of asymmetric preferences is loss function sensitive and may lead to incorrect inference if the loss function is misspecified, whereas our test can be used without requiring the specification of a particular functional form for the loss. In addition, simulations show that our proposed test has good finite sample properties even when the forecasting rule is nonlinear.

Our empirical study highlights some differences in the results that we obtain when applying the two forecast rationality tests to data from the Survey of Professional Forecasters. The contradiction in the results reveal interesting insights regarding the rationality of the SPF forecasts.

Chapter 2

Global Identification Failure in DSGE Models and its Impact on Forecasting

2.1 Introduction

Dynamic stochastic general equilibrium (DSGE) models use a coherent theoretical framework to understand and explain macroeconomic fluctuations. Important policies are routinely produced based on the estimates of DSGE parameters. Before performing an empirical analysis with a particular model, it is essential however that the model be checked for identifiability. Assessing the identifiability of DSGE models is important for both econometric rigour as well as for delivering reliable policy recommendations. On the one hand, a parameter that is not identifiable cannot be consistently estimated. Identification issues can lead to misleading estimation and inference. On the other hand, as the policy implications of observationally equivalent parameter points

can be distinct, the reliability of policy recommendations depends upon the primitive assumption of identifiability.

Although the validity of statistical inference in DSGE models has been questioned due to identification deficiencies (see for e.g. Beyer and Farmer [2004], Canova and Sala [2009]), there are relatively few papers in the literature that assess the impact of identification failure on different macroeconomic outcomes [1]. The purpose of this paper is to examine whether the fragility of parameter estimates in DSGE models, resulted from *lack of identification*, transfers to other objects of interest such as point forecasts. A few papers in the literature analyzed the effect of identification loss on impulse response functions (see for e.g. Morris [2014]), but to my knowledge no existing paper examines the impact of identification failure on forecasting with DSGE models.

In this paper, we focus on *global identification failure*, as this particular identification issue is of interest when giving policy recommendations based on DSGE models. Other types of identification failures such as local and weak identification, even though interesting from the point of view of the econometric challenges that they imply, they are less relevant from a policy perspective. We illustrate the results through the model of An and Schorfheide [2007], a small scale DSGE model, representative for the models currently used in monetary policy analysis at policy institutions. This particular model has been chosen for the analysis motivated by the fact that the dimension of the parameter space is not too large and thus, it permits the computation of the numerical results in reasonable times.

As there are currently no available analytical results for global identification in DSGE models, we proceed by assessing global identifiability numer-

[1] Such as impulse responses, forecasts, etc.

ically. To this end, we construct a fine grid on the economically meaningful parameter space, for the starting values used in the optimization of the objective function. Using these starting values, the aim is to find through a global optimization algorithm, several *distinct* maximizing parameter values yielding the *same* value for the likelihood. In searching for observationally equivalent points, we treat two cases: optimizing the log-likelihood (frequentist approach) and optimizing the log-posterior (Bayesian approach).

We show numerically that when the model is solved with linear approximations, three parameters: the inverse elasticity of demand (ν), the price stickiness (ϕ), and the steady state of government spending ($1/g$) fail to be globally identified. In contrast, under nonlinear approximations, where second-order accurate solutions to the DSGE model are obtained from a second order expansion of the equilibrium conditions, these three parameters become globally identifiable. This sheds some light on a scarcely investigated issue in the literature: the log-linearization error in DSGE identification. The reason why in this particular model, quadratic approximations can identify the structural parameters that are not identifiable under the log-linearized model is that ν and ϕ become linearly independent in the nonlinear version of the Philips curve, and g *does* appear in the nonlinear equilibrium equations, while it did not enter the linear likelihood. These findings confirm the hypotheses made in An and Schorfheide [2007], based on likelihood contour graphs.

Based on observationally equivalent parameter points obtained from the grid search results, we perform several forecasting exercises: i) forecasting with observationally equivalent parameter values from a correctly specified model, ii) forecasting with observationally equivalent parameter values from misspecified models and iii) forecasting with observational equivalent parameter points

arising from different model structures. The focus here is on the linear model only, where there is evidence for global identification failure.

The results show that when observationally equivalent parameter estimates belong to the same model structure, they imply the same system matrices, which result in the same point forecasts, under both correct and misspecification. Thus, global identification failure does not impose issues in constructing predictions from DSGE models, when identification is conditional on the model structure, i.e. in a framework where there could exist different parameter values *within the same* DSGE structure that lead to the same dynamics of the observables. However, when the identification problem is analyzed permitting different structures, (e.g. models with different policy rules, different types of frictions or shock processes), two parameter points that generate the same data dynamics can lead to distinct forecasts. This is illustrated with observationally equivalent parameter estimates arising from two model structures: one in which the monetary policy follows an interest rate rule that reacts to current inflation, and another where the central bank responds to expected inflation. The differences in the obtained forecasts can be large. With a current inflation specification, the forecasted quarter-to-quarter GDP growth rates are around -0.5% and the annualized quarter-to-quarter inflation rates are around 2.8%, while with an expected inflation specification, the quarter-to-quarter GDP growth rates are around 1.9% and the annualized quarter-to-quarter inflation rates are around 5%.

2.1.1 Related Literature

The paper is related to two main strands of literature. The first line concerns the literature on DSGE identification - which is by now fast growing. One of the first papers that draw attention to the identification problems in these models was Canova and Sala [2009]. They suggest several diagnostic procedures for identification based on graphical and simulation analyses.

A couple of papers made advances in providing formal conditions for *local identification*. Iskrev [2010] obtains sufficient (but not necessary) conditions for local identification based on the rank of the Jacobian matrix that maps the deep parameters to first and second moments of the observables. His approach is suitable when the DSGE model is estimated by likelihood based methods. Komunjer and Ng [2011] establish rank and order conditions for local identification from restrictions implied by equivalent spectral densities for both singular [2] and non-singular models. Their established conditions are independent of the data or the estimator used and can be verified prior to estimation. Qu and Tkachenko [2012] obtain necessary and sufficient condition for local identification formulated in the frequency domain. Related to nonlinear DSGE models, Morris and Lee [2014] provide rank and order conditions for local identification of parameters in nonlinear models.

Another stream of literature focuses on the issue of *global identification*. Fukač, Waggoner, and Zha [2007] study global identification in invertible DSGE models, with solutions in a VAR form. Qu and Tkachenko [2013] introduce a criterion based on the Kullback-Leibler distance between two DSGE models computed in the frequency domain and shows that global identification

[2]Singularity occurs when there are fewer shocks than observables.

fails if and only if the criterion equals zero when minimized over the relevant region of the parameter space. They consider log-linearized DSGE models, and treat both determinacy and indeterminacy [3] cases, the results being applicable for singular and non-singular models. Basing their analysis on the time domain, Kociecki and Kolasa [2014] obtain a necessary (order) condition for global identification and develop an efficient algorithm that checks for global identification by shrinking the space of admissible parameter values. The last two approaches for checking global identification in DSGE models are (computationally intensive) numerical methods and the literature still lacks complete formal analytical conditions for global identification. The above mentioned papers examine identification of the DSGE system as a whole. An earlier strand of the literature raised questions about identification of particular equations in a DSGE model such as the Philips curve: Mavroeidis [2005], Kleibergen and Mavroeidis [2009] or the Taylor rule: Cochrane [2011]

Related to the problem of *weak identification* [4] in DSGE models, several inference methods robust to weak identification have been proposed. These methods are all based on the inversion of test statistics. Andrews and Mikusheva [2014] focus on testing and confidence set construction using two weak identification-robust forms of the classical LM test. Working in the frequency domain, Qu [2014] develops asymptotically valid confidence sets for the struc-

[3] Indeterminacy appears when the log-linearized models has multiple solutions. In this case, the state-space representation is not uniquely determined by the model because the variables that enter the state transitions equation as well as the matrices that appear in the transition equation differ according to the solution of the model.

[4] When a DSGE model is weakly identified, the likelihood function is not completely flat, but has very low curvature with respect to some parameters. Weak identification causes statistics to be non-normally distributed even in large samples. Also, as illustrated in Guerron-Quintana et al. [2013], the usual asymptotic equivalence results between Bayesian and frequentist inference are no longer valid - the Bernstein-von Mises theorem does not apply. As a result, under weak identification, the posterior mean of the posterior distribution is not a consistent estimator.

tural parameters and uniform confidence bands for the impulse response functions. Dufour, Khalaf, and Kichian [2013] propose a limited information approach and suggest the inversion of moment based tests, the approach being applicable to DSGE models that have a finite order VAR representation. Guerron-Quintana, Inoue, and Kilian [2013] construct confidence sets that remain asymptotically valid when the structural parameters are weakly identified by inverting the likelihood ratio statistic and the Bayes factor. A test that detects weak identification is proposed in Inoue and Rossi [2011]. They implement their approach to test the null of identification against weak (or no) identification of the structural parameters in a Taylor rule monetary policy reaction function.

More recently, the literature started to exploit *time variation in parameters* as a source of identification. Magnusson and Mavroeidis [2014] show that time variation in a subset of parameters can be used in the identification of parameters that are stable overt time. Canova et al. [2015] illustrate that identification problems in DSGE models may arise as a result of misspecification due to neglected parameter variation. In a related literature that focuses on testing the *correct specification* of DSGE models, Komunjer and Zhu [2016], highlight that the parameters (system matrices) of the DSGE's state space model representation are not identified, and construct a likelihood ratio test, using classical testing theory, for the validity of the DSGE model, taking into account this lack of identification.

The second main strand of literature this paper is related to, is the literature on forecasting with DSGE models. These models have become by now part of the set of models used for forecasting in central banks and policy

institutions. The particular interest in forecasting with theoretically-grounded DSGE models among policy institutions, is motivated by the fact that compared to a-theoretical models (such as VARs, BVARs or DFM), DSGE models have a high degree of theoretical coherence, they deliver a richer interpretation of the results, and can be used to construct predictions on the effects of alternative policy scenarios.

The forecasting performance of DSGE models has been shown to be competitive with alternative prediction methods used at central banks. Results from the seminal papers Smets and Wouters [2003] and Smets and Wouters [2007] suggest that DSGE model based forecasts compare well with forecasts obtained from reduced form models such as standard and Bayesian VARs. Similar results have been obtained by Edge, Kiley, and Laforte [2010] comparing the FRB's DSGE model to alternative reduce-form time series models, the Greenbook forecasts and the FRB/US model forecasts. The forecasting performance of DSGE models for the Euro area has been assessed by Christoffel, Coenen, and Warne [2010] (paper uses the NAWM model developed at the ECB) and Adolfson, Lindé, and Villani [2007b]. The forecasting accuracy of the Riksbank's DSGE model has been evaluated by Adolfson et al. [2007a].

All these studies differ in terms of the estimation sample, forecasting sample, variable definitions or data vintages used. In order to make the DSGE forecasting performance results comparable across different studies, in a survey paper, Del Negro and Schorfheide [2013] compute the ratio between the RMSE reported in various papers and the RMSE of a benchmark AR(2) model. Overall, they find that in terms of forecast accuracy, DSGE models compare well with alternative prediction methods - especially for the output growth and especially for medium-run forecasts (four quarter ahead horizon).

Nevertheless, some authors bring critiques on the forecasting performance of DSGE models (see for e.g. Giacomini [2014], Gurkaynak et al. [2013], Giacomini and Rossi [2012]). They emphasize that the conclusion reached by the aforementioned studies is sensitive to arbitrary choices that one makes (such as the choice of the observables, the choice of the priors and hyperparameters, data processing such as detrending, or how one deals with stochastic singularity), when assessing the forecasting performance of DSGE models.

Beyond forecasting with standard DSGE models, another strand of literature, focuses on obtaining forecasts combining DSGE models with VARs (see e.g. Del Negro and Schorfheide [2004], Amisano and Geweke [2013], Waggoner and Zha [2012]) or with DFM (see e.g. Schorfheide, Sill, and Kryshko [2010]; incorporating external information into DSGE model based forecasts such as nowcasts, expectations on inflation, output and interest rate (Del Negro and Schorfheide [2013]); or incorporating theoretical restrictions into a-theoretical models using exponential tilting (Giacomini and Ragusa [2014]).

Section 2.2 sets up the econometric framework and presents the specificities of the identification problem in DSGE models. Section 2.3 illustrates numerically the failure of global identification in a small scale prototypical DSGE model. The model is analyzed in both log-linear and nonlinear form. Several forecasting exercises based on observationally equivalent parameter points are performed in Section 2.4. Section 2.5 concludes.

2.2 The Identification Problem in DSGE Models

In this section, we introduce the ABCD representation of a DSGE model, present the particularities of the model that make the classical identification results inappropriate in this framework and highlight the implications of identification failure in both the classical and Bayesian estimation approaches.

2.2.1 Setup

Consider a DSGE model with structural parameters θ belonging to a set $\Theta \subseteq \mathbb{R}^{n_\theta}$. Let X_t be an $n_X \times 1$ state vector, Y_t an $n_Y \times 1$ vector of observables and ϵ_t a vector of structural shocks of size $n_\epsilon \times 1$. After writing down the system of nonlinear expectational equations of the model, a solution to it can be obtained, which will have the form given in (2.1).

$$X_t = f(X_{t-1}, \epsilon_t, \theta) \tag{2.1}$$

The state transition equation (2.1) is pinned down by either finding a close form expression for the function $f(\cdot)$, or more frequently, by numerical approximation [5]. The majority of the literature focuses however on the log-linearized model around the steady-state, in order to work with a linear version of the state transition equation:

$$X_t = AX_{t-1} + B\epsilon_t \tag{2.2}$$

[5]Several solution algorithm are available, such as: Uhlig et al. [1999], Klein [2000], Christiano [2002], Sims [2002], King and Watson [2002], Schmitt-Grohé and Uribe [2004].

where the matrices A and B are nonlinear transformations of the structural parameter θ. In order to compute a likelihood for the DSGE model, a set of observable variables Y_t has to be chosen and a measurement equation such as (2.3) specified.

$$Y_t = CX_{t-1} + D\epsilon_t \tag{2.3}$$

The matrices C and D depend nonlinearly on θ. Equations (2.2) and (2.3) form the state-space representation or (ABCD representation) of the model:

$$\begin{aligned} \underbrace{X_t}_{n_X \times 1} &= \underbrace{A}_{n_X \times n_X} \underbrace{X_{t-1}}_{n_X \times 1} + \underbrace{B}_{n_X \times n_\epsilon} \underbrace{\epsilon_t}_{n_\epsilon \times 1} \\ \underbrace{Y_t}_{n_Y \times 1} &= \underbrace{C}_{n_Y \times n_X} \underbrace{X_{t-1}}_{n_X \times 1} + \underbrace{D}_{n_Y \times n_\epsilon} \underbrace{\epsilon_t}_{n_\epsilon \times 1} \end{aligned} \tag{2.4}$$

If the state-space representation is linear such as (2.4), it can be estimated using the Kalman filter. For nonlinear state space models, typically the particle filter (see for e.g. Pitt and Shephard [1999]) is used. The ABCD representation of the model constitutes the starting point of most of the analysis in the literature and has a central roll when discussing identification in DSGE models.

2.2.2 Definitions

Identification problems in DSGE models arise when different values of the structural parameters lead to the same joint distribution of the observables. In other words, changes in some of the parameters may result in indistinguishable outcomes.

To fix ideas, it is worthwhile to recall at this point the following defini-

tions related to the problem of identifiability in the context of DSGE models, (Rothenberg [1971]).

Definition 1. Given a $n_Y \times 1$ vector of observed data Y, a $n_\theta \times 1$ vector of DSGE structural parameters of interest $\theta \in \Theta \subseteq \mathbb{R}^{n_\theta}$ and the likelihood function $\mathcal{L}(\theta; Y)$, two parameter points of a DSGE model θ_0 and θ_1 belonging to the set $\Theta \subseteq \mathbb{R}^{n_\theta}$ are said to be *observationally equivalent* if $\mathcal{L}(\theta_0; Y) = \mathcal{L}(\theta_1; Y)$ for all Y.

Remark 3. Note that the concept of observational equivalence is not restricted to a particular sample. In other words, two points are observationally equivalent if the likelihood function evaluated at these two points has the same value, regardless of the data set used.

This work considers however *sample* observational equivalence, because in practice, when a researcher estimates a DSGE model and faces the problem of identification failure, she works with a particular data set.

Remark 4. In a Bayesian setup too, observational equivalence reduces to the equality of the likelihoods.

Definition 2. A DSGE model is *locally identifiable* at a point $\theta_0 \in \Theta$ if there exist an open neighborhood of θ_0 such that for every θ_1 in this neighborhood, θ_0 and θ_1 are observationally equivalent if and only if $\theta_0 = \theta_1$.

Definition 2 states that a parameter point is locally identified when it is uniquely distinguishable in an ϵ-neighbourhood. Local identification is a necessary condition for well behaved estimators to exist, in the sense of having a well behaved distribution. Establishing local identifiability is useful if it is a step towards global identifiability, and in many situations it is not the final

product of interest. As pointed out in Komunjer and Ng [2011], even if a model is locally identified at every point in the parameter space, it can still fail to be globally identified.

The main focus of this paper is on global identification.

Definition 3. A DSGE model is *globally identifiable* at a point $\theta_0 \in \Theta$ if there is no other θ_1 in the parameter space Θ that is observationally equivalent to θ_0.

Global identification deals with the question of whether there exists another point in the parameter space that results in the same macroeconomic dynamics: autocovariances, impulse responses. This is a question of interest for economists, because if the model is not globally identified, there exist multiple likelihood maximizing parameter values in the admissible parameter space, which can have important consequences on quantitative policy analysis, forecasting or scenario simulations. For example, a recent work by Morris [2014], documents the effect of a multiplicatively-valued maximum likelihood estimator on monetary policy impulse responses. He shows that observationally equivalent parameter points can lead to distinct monetary policy impulse responses and thus yield different macroeconomic consequences.

2.2.3 Non-identifiability of the ABCD matrices

Finding analytical conditions for identification in DSGE models has proven to be a difficult problem for several reasons. First of all, compared to linear simultaneous equations, identification in DSGE models is less transparent because the system matrices of the state-space representation are highly nonlinear functions of the structural parameters and they can only be evalu-

ated numerically. Secondly, the classical identification results for simultaneous equations such as Fisher [1966], Koopmans [1950], Rothenberg [1971] are valid only for static models, while DSGE models are dynamic by construction. Perhaps the most important reason why the classical identification results fail in the case of DSGE models is that the rank conditions from Rothenberg [1971] [6] are based on the assumption that the reduced form parameters are identifiable, an assumption that is not verified in the case of DSGE models. Thus, the classical approach of identifying structural parameters from reduce form parameters cannot be applied as the state-space solution is not a standard reduced form. To see this, consider the following result that characterizes observational equivalence in singular systems.

Proposition 1. *(Komunjer and Ng [2011]): Observational Equivalence,* $n_\epsilon \leq n_Y$

Under Assumptions 1, 2, 4S and 5S from Komunjer and Ng [2011], θ_0 *and* θ_1 *are said to be observationally equivalent if and only if there exists a full rank* $n_X \times n_X$ *matrix* T *and a full rank* $n_\epsilon \times n_\epsilon$ *matrix* U *such that:*

$$\begin{aligned} A(\theta_1) &= TA(\theta_0)T^{-1} \\ B(\theta_1) &= TB(\theta_0)U \\ C(\theta_1) &= C(\theta_0)T^{-1} \\ D(\theta_1) &= D(\theta_0)U \\ \Sigma_\epsilon(\theta_1) &= U^{-1}\Sigma_\epsilon(\theta_0)U^{-1'}. \end{aligned} \tag{2.5}$$

[6] Rothenberg [1971] shows that a parameter is identified at a point θ_0, if the information matrix evaluated at θ_0 is non-singular.

Proposition (1) is used to derive formal identification (rank and order) conditions for the structural parameters θ. The immediate implication of Proposition (1) is that the ABCD matrices are not identified. To clarify this, start from the state space representation of the model:

$$X_t = A(\theta)X_{t-1} + B(\theta)\epsilon_t$$
$$Y_t = C(\theta)X_{t-1} + D(\theta)\epsilon_t$$

Multiply by T the transition equation to get:

$$TX_t = TA(\theta)X_{t-1} + TB(\theta)\epsilon_t$$

which is equivalent to

$$TX_t = TA(\theta)T^{-1}TX_{t-1} + TB(\theta)UU^{-1}\epsilon_t$$

that can be written as

$$\underbrace{TX_t}_{\widetilde{X_t}} = \underbrace{TA(\theta)T^{-1}}_{\widetilde{A(\theta)}}\underbrace{TX_{t-1}}_{\widetilde{X_{t-1}}} + \underbrace{TB(\theta)U}_{\widetilde{B(\theta)}}\underbrace{U^{-1}\epsilon_t}_{\widetilde{\epsilon_t}}$$

Analogously, the measurement equation can be written as:

$$Y_t = C(\theta)T^{-1}TX_{t-1} + D(\theta)UU^{-1}\epsilon_t$$

or

$$Y_t = \underbrace{C(\theta)T^{-1}}_{\widetilde{C(\theta)}}\underbrace{TX_{t-1}}_{\widetilde{X_t}} + \underbrace{D(\theta)U}_{\widetilde{D(\theta)}}\underbrace{U^{-1}\epsilon_t}_{\widetilde{\epsilon_t}}$$

The quadruples $(A(\theta), B(\theta), C(\theta), D(\theta))$ and $(\widetilde{A(\theta)}, \widetilde{B(\theta)}, \widetilde{C(\theta)}, \widetilde{D(\theta)})$ are said to be related by a similarity transformation. Similarity transforms can rotate the latent variables $(X_t,\ \epsilon_t)$ of the state space model and leave unchanged the autocovariances [7] of the observables Y_t. Consequently, the information contained only in the autocovariances of Y_t is not sufficient to identify the state-space model parameters.

2.2.4 Implications of Identification Failure in Frequentist and Bayesian Analysis

There are several approaches considered in the literature for the estimation of DSGE models: maximum likelihood, GMM, minimum distance estimation, Bayesian estimation. Regardless of using a classical inferential framework or Bayesian methods, all estimation techniques can be affected by the possible lack of identification of the parameters of interest. In a classical framework, as it is well known, standard estimation and inference technique become unreliable, if the parameters are not identified. The reason is that these estimation methods, are based on the existence of a *unique* true parameter value that satisfies the model's moment conditions or minimizes a loss function. If this condition is not satisfied, the estimators are in general inconsistent, have non-standard distributions and require adjusted inference techniques for hypothesis testing. The validity of confidence interval coverage is usually also affected. Moreover, a technical issue that appears in the classical approach is that the numerical optimization methods may have convergence issues in the case of a relatively flat region of the objective function, which is typical when some

[7] Komunjer and Ng [2011] discusses identifiability from $\Gamma_Y(j, \theta_0)$ - the autocovariance matrix of Y at each j.

parameters are only weakly identified.

While the consequences of identification failure in a classical inference framework are clear, they may not be evident when Bayesian methods are used to estimate a DSGE model[8]. However, it is important to account for identifiability, even if one uses a Bayesian approach. Although identification is not required as a regularity condition for the Bayesian inference, lack of identification may complicate the generation of draws from the posterior distribution because if a parameter is not identified, there will be no learning from the data and the posterior will only reflect prior information (see for e.g. Poirier [1998], Müller [2012], Koop, Pesaran, and Smith [2013]).

When DSGE models are estimated by posterior simulation algorithms, checking for identification is commonly accomplished by comparing posteriors to priors, motivated by the fact that when some parameter component is not identified, the likelihood function is flat in that direction of the parameter space and the prior will not be updated. However, checking whether the posterior differs from the prior, to check the informativeness of the data, is not sufficient to reveal identification problems as posteriors can differ from priors even for unidentified parameters. As argued in Poirier [1998], Müller [2012], Koop, Pesaran, and Smith [2013], when the unidentified parameter components are a priori correlated with the identified ones, the posterior of an *unidentified* parameter component can substantially differ from its prior. In this case, the researcher cannot distinguish whether learning occurs from the data or from the dependence in the prior (see for example *Proposition 2* in Poirier [1998] or Figure 1 and Figure 3 in Koop, Pesaran, and Smith [2013]). Thus, Bayesian

[8] An argument commonly used in the literature to support a Bayesian setting is that as long as the prior distribution is proper, the posterior distribution will be well defined, regardless of identification problems.

methods and in particular this informal check of identification may conceal the identification problems in DSGE models.[9]

2.3 Global Identification Failure: an Illustration

In this section, we illustrate numerically the failure of global identifiability of structural parameters in the model of An and Schorfheide [2007]. This is accomplished by performing a grid search on an economically admissible parameter space. The analysis is carried out on the model solved with both linear and nonlinear approximations.

2.3.1 The Model of An and Schorfheide (2007)

The model of An and Schorfheide [2007] is a small scale DSGE model, representative for the models currently used in monetary policy analysis at policy institutions. This particular model has been chosen for the analysis given that the dimension of the parameter space is not too large and the numerical results can be obtained in reasonable times.

After log-linearizing the model's equilibrium conditions around the steady

[9] A Bayesian *local* identification indicator that does not suffer from this drawback was recently proposed in Koop, Pesaran, and Smith [2013]. The indicator is based on the idea that for identified parameters, the posterior precision rises with the sample size T, while for unidentified parameters the posterior precision might be updated, but its rate of update is slower than T.

state, the following system of linear expectational equations is obtained:

$$\begin{aligned} x_t &= E_t x_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tau}(r_t - E_t \pi_{t+1} - E_t z_{t+1}) \\ \pi_t &= \beta E_t \pi_{t+1} + \kappa(x_t - g_t) \\ r_t &= \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2(x_t - g_t) + \epsilon_{rt} \\ g_t &= \rho_g g_{t-1} + \epsilon_{gt} \\ z_t &= \rho_z z_{t-1} + \epsilon_{zt} \end{aligned} \tag{2.6}$$

where $\kappa = \frac{\tau(1-\nu)}{\nu\pi^2\phi}$. The variables x_t, π_t, r_t, g_t and z_t stand for output, inflation, nominal interest rate, government spending and disturbance in the technology growth, respectively. The endogenous variables are driven by three mutually uncorrelated structural shocks: a monetary policy shock, $\epsilon_{rt} \sim N(0, \sigma_r^2)$, a government spending shock, $\epsilon_{gt} \sim N(0, \sigma_g^2)$ and a technology growth shock, $\epsilon_{zt} \sim N(0, \sigma_z^2)$. The system is square as the number of structural shocks is equal to the number of observables.

The vector of observables is $y_t = (x_t, \pi_t, r_t)'$ and the vector of state variables is $s_t = (\pi_t, x_t, r_t, g_t, z_t, E_t\pi_{t+1}, E_t x_{t+1})'$. The linear solution of the DSGE model, for the vector of model parameters, θ is given by

$$s_t = \Phi_1(\theta) s_{t-1} + \Phi_2(\theta)\epsilon_t \tag{2.7}$$

where $\epsilon_t = (\epsilon_{zt}, \epsilon_{gt}, \epsilon_{rt})'$ is the vector of structural shocks. In the paper, standard algorithm such as Sims [2002] is used to obtain the first-order accurate solution of the DSGE model and the algorithm of Schmitt-Grohé and Uribe [2007] is used for the second-order accurate solutions. The model is completed by defining a set of measurement equations that relate the elements of s_t to

the set of observables y_t.

$$y_t = \Psi_1(\theta) + \Psi_2(\theta)s_t \tag{2.8}$$

In the numerical simulations of the paper the time period corresponds to one quarter and the observations available for estimation are: quarter-to-quarter per capita GDP growth rates (XGR), annualized quarter-to quarter inflation (INF) and annualized nominal interest rates (INT). The three series are measured in percentages and are related to the model variables as follows:

$$XGR_t = \gamma^Q + 100(x_t - x_{t-1} + z_t)$$
$$INF_t = \pi^A + 400\pi_t$$
$$INT_t = \pi^A + r^A + 4\gamma^Q + 400r_t$$

where the parameters γ^Q, π^A and r^A are related to steady states of the model economy as:

$$\gamma = 1 + \frac{\gamma^Q}{100}$$
$$\beta = \frac{1}{1 + r^A/400}$$
$$\pi = 1 + \frac{\pi^A}{400}$$

The transition equation given in 2.7 and the measurement equation given in 2.8 form the state-space representation of the linearized model. The system matrices Φ_1, Φ_2, Ψ_1 and Ψ_2 are of size $(n_s \times n_s)$, $(n_s \times n_\epsilon)$, $(n_y \times 1)$ and $(n_y \times n_s)$, respectively, where n_s, n_ϵ and n_y are the sizes of the vectors s_t, ϵ_t and y_t. The elements of these matrices are nonlinear functions of the DSGE

structural parameters, θ. Writing the DSGE model in a linear state-space form is convenient because under the assumption that the innovations ϵ_t are Gaussian, Kalman filtering techniques can be used to evaluate the likelihood function. Most of the literature works with the log-linearized model in order to obtain a linear version of the state transition equation.

The structural parameters of the linearized DSGE model are collected in a 13-dimensional vector,

$$\theta_1 = [\tau, \kappa, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, r^A, \pi^A, \gamma^Q, \sigma_r, \sigma_g, \sigma_z][10].$$

The description for each of these parameters can be found in Table 2.1. Note that the linearized model is expressed in terms of the composite parameter κ because the inverse elasticity of demand (or degree of imperfect competition) ν and the price stickiness ϕ do not enter the linear likelihood and they are not separately identifiable in the linear version of the An and Schorfheide [2007] DSGE model .

In the case of a nonlinear DSGE model, quadratic approximations are used for the model's solution and the likelihood can be evaluated by the particle filter. With nonlinear approximations, the vector of structural parameters for this particular model has 15 elements:

$$\theta_2 = [\tau, (\nu, \phi), 1/g, \psi_1, \psi_2, \rho_r, \rho_g, \rho_z, r^A, \pi^A, \gamma^Q, \sigma_r, \sigma_g, \sigma_z].$$

Here, the reduced form Phillips curve parameter (composite parameter) κ has been replaced by the couple (ν, ϕ) [11] and the parameter vector is augmented

[10] Note that $r = \gamma/\beta$.

[11] Alternatively, k can be replaced by (ν, κ) too, keeping in mind that $\kappa = \frac{\tau(1-\nu)}{\nu\pi^2\phi}$.

with $1/g$ which is equal to the steady state ratio c/y. The paper by An and Schorfheide [2007] provides a graphical intuition on why the parameters ν and ϕ might become separately identifiable under quadratic approximations. We show numerically in Section 2.3.4 that under nonlinear approximations, (ν and ϕ) do become separately identifiable. We also show that the steady state of government spending g, which is not identifiable under linear approximations, as it appears in the second-order approximations only, is indeed identified in a nonlinear DSGE model.

2.3.2 Searching for Observationally Equivalent Parameter Points

As there are currently no available analytical results for verifying global identification in DSGE models, we proceed by checking numerically whether we can find observationally equivalent parameter points in an economically admissible parameter space. To this end, we construct a fine grid on the economically meaningful parameter space, for the starting values in the optimization of the objective function. The aim of this search is to find several *different* maximizing parameter values yielding the *same* value of the likelihood, for one simulated data set. As previously mentioned, for the purpose of this paper it is enough to find sample observationally equivalent points, as in a practical application, one will work with a particular data set.

The grid is constructed as follows. First, for each component of θ, some bounds [12] are fixed as given in Table 2.1 and 100 equally spaced points are selected on each interval to make the grid fine enough. How fine the grid is

[12] These bounds are fixed based on an economically admissible parameter space.

along each dimension will depend on the actual length of the fixed intervals, but this does not impact the results. Then, the coordinates of a particular point on the grid (point that will constitute the starting value for the optimization of the objective function) are obtained by selecting a uniform random draw from the 100 fixed ones, along each dimension of the parameter space. The grid that is eventually obtained is 13-dimensional in the case of the linear model and 15-dimensional in the case of the nonlinear model.

Table 2.1: Grid Search Parameter Bounds

Parameter	**Description**	**Bounds**
τ	CRRA	[0.1000, 4.0000]
κ	Composite parameter	[0.1000, 0.4000]
ψ_1	Taylor rule infl. coeff.	[1.0000, 2.0000]
ψ_2	Taylor rule out. coeff.	[0.0001, 0.8000]
ρ_r	r_t persistence	[0.4000, 0.9000]
ρ_g	g_t persistence	[0.8000, 0.9900]
ρ_z	z_t persistence	[0.8000, 0.9900]
r^A	St. st. interest rate	[0.5000, 1.5000]
π^A	St. st. inflation	[2.0000, 4.0000]
γ^Q	St. st. Av. output growth rate	[0.1000, 0.9000]
$100\sigma_r$	ϵ_{rt} std. err.	[0.0001, 0.0050]
$100\sigma_g$	ϵ_{gt} std. err.	[0.0001, 0.0090]
$100\sigma_z$	ϵ_{zt} std. err.	[0.0001, 0.0060]
ϕ	Index of price stickiness	[30.000, 90.000]
ν	Inv. elast. of demand	[0.0100, 0.2000]
$1/g$	g - St. st. gov. spending	[0.8000, 0.9500]

NOTE: The table reports the Parameter Bounds used in the Grid Search for each of the structural parameters of the An and Schorfheide model. Recall that ν and $1/g$ only affect the second order accurate solution of the DSGE model.

The optimization of the objective function is performed using a quasi-Newton algorithm available in the Matlab function "CSMINWEL" by Sims. It is a minimization routine - the negative likelihood is minimized. The routine finds the global minimum, being robust to certain pathologies of the likelihood such as "cliffs".

In searching for observationally equivalent points, we treat two cases:

optimizing the log-likelihood (frequentist approach) and optimizing the log-posterior (Bayesian approach). When the grid is constructed for the initial values of the log-posterior optimization, we make use of the prior distributions from An and Schorfheide [2007]. They are reproduced for convenience in Table B.1 and B.3.

2.3.3 Numerical Evidence: Linear Model

Departing from the hypotheses made in An and Schorfheide [2007], namely that in the linear approximation of the model, the inverse elasticity of demand (ν) and the price stickiness (ϕ) are not *separately* identifiable, and that the steady state of government spending ($1/g$) is *unidentified*, we estimate the linear version of the DSGE model fixing all other parameters to their true values (see Table B.1 for the DGP parameter values) and assess the global identifiability of these three parameters. The model is estimated from simulated data of 80 observations generated from the same model (correct specification).

Optimizing the log-likelihood (Frequentist approach). Table B.5 reports the grid search results for the case in which the objective function that is optimized is the log-likelihood. Over 100 iterations have been made in the grid search, but here we report the first 40 iterations only. Columns *Init* ν, *Init* ϕ and *Init* $1/g$ list the starting grid points used for the optimization of the log-likelihood. The location of the obtained global extremum in each iteration is given in the next three columns. Columns *LogPost*, *LogLike* and *LogPrior* give the value of the log-posterior, log-likelihood and log-prior evaluated at the obtained mode. Recall that two parameter points of a DSGE model θ_0 and θ_1 are observationally equivalent if $\mathcal{L}(\theta_0; Y) = \mathcal{L}(\theta_1; Y)$. The grid search

results show that observationally equivalence is prevalent, several observationally equivalent parameter points are found. For different parameter values, we get back the same value of the likelihood, indicating the failure of global identifiability of these three parameters and confirming the hypotheses made in An and Schorfheide [2007]. Some parameter points (in italics) could be eliminated from the estimation problem as they are out of the economically admissible parameter space, but even without them, we are still left with plenty of observationally equivalent points that would be hard for the researcher to chose from.

Optimizing the log-posterior (Bayesian approach). The grid search results for the optimization of the log-posterior are reported in Table B.6. As previously, 40 iterations are reported. After the global extremum of the log-posterior is found, it is plugged back in the log-likelihood (Column *LogLike*). It is interesting to notice that the variation in the parameter points (Columns ν, *phi* and $1/g$) is much more reduced compared to the results obtained when the the log-likelihood was optimized. Nevertheless, even when one makes use of some prior information, it is clear that ν, ϕ, $1/g$ are not globally identified - take for example parameter points with ID: 2, 9 and 21 which are distinct but all have a log-likelihood value of -208.74. Also, the use of prior information helps in eliminating some of the economically implausible estimates, that could appear when the maximum likelihood is used. Note however that not all parameter points that are out of the economically plausible parameter space are eliminated when using prior information and optimizing the log-posterior - see for example lines with ID: 17 and 39, in italics.

Next, in Table B.7, we report results from a grid search setting that

confirms that all parameters with the exception of ν, ϕ and $1/g$ are identified in the linear model. In this example, I estimate a model with $\theta = [\tau, (\nu, k), 1/g, \psi_1, \psi_2, \rho_r, \rho_g,$
$\rho_z, r^A, \pi^A, \gamma^Q, \sigma_r, \sigma_g, \sigma_z]$[13]. The parameters ν and $1/g$ are fixed to their true values, 0.1 and 0.85, respectively. Instead of ϕ we have now the composite parameter k, which is identified and thus estimated. In the setting of Table B.7, the model is estimated using an output growth specification in the monetary policy rule, i.e.

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 \pi_t + (1-\rho_r)\psi_2(\Delta x_t + z_t) + \epsilon_{rt} \tag{2.9}$$

on data generated from the same model (thus correct specification). The objective function that is optimizes is the log-posterior.

The results of Table B.7 confirm that all parameters that are not fixed are indeed identified. What is revealed however is that in this case, the log-posterior (the log-likelihood as well) of the model has two modes of different height: -144.86 and -138.50, leading thus to two sets of parameter estimates. This problem can arise frequently in practice. It does not constitute however an identification loss. It is rather an issue of multiple modes.

So far, the grid search was performed for models that are correctly specified. Table B.8 and B.9 report now results for two misspecification cases. Data is generated using an output gap specification in the monetary policy rule (see the interest rate equation in 2.6), while estimation is performed first under an output growth specification as given by equation 2.9 and then under

[13]See footnote 11 for details on why (ν, ϕ) can be replaced by (ν, k).

an expected inflation specification as in equation 2.10 below:

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 E_t \pi_{t+1} + (1-\rho_r)\psi_2 (x_t - g_t) + \epsilon_{rt} \tag{2.10}$$

In the first misspecification case, the monetary policy follows an interest rate rule that reacts to the output growth (Table B.8), while in the second case the central bank responds to expected inflation rather than current inflation (Table B.9).

While these last two grid search results in the linear model case, do not present further novelty in terms of parameter identification, they are needed in Section 2.4, where the forecasting exercise is performed.

2.3.4 Numerical Evidence: Nonlinear Model

As pointed out in Giacomini [2013], an issue that is scarcely investigated in the literature, is the effect of log-linearization on the identification of structural parameters. In this section, we perform a grid search analysis on the DSGE model solved with second-order approximations and provide some useful evidence towards the importance of the log-linearization error in identification.

The grid search in the previous section has been performed on the model solved with linear approximations. Linear approximation methods are widely used in the empirical literature because they lead to a state-space representation of the model that can be analyzed via the Kalman filter. Here, we now turn to nonlinear approximations, where second-order accurate solutions to the DSGE model are obtained from a second order expansion of the equilibrium conditions. When a DSGE model is solved with second-order approximations, the linear Kalman filter cannot be used to compute the likelihood function. We

follow the standard approach to evaluate the likelihood function in this case, by using a particle filter (Fernández-Villaverde and Rubio-Ramírez [2007]). Nonlinear DSGE models are relevant when the goal is to analyze asset pricing implications of DSGE models, or to compare welfare across policy regimes.

The grid search results on the model solved with second-order approximations are reported in Table B.10 where the sample size is N=80 and Table B.11 where the sample size of N=3000. None of the parameters are fixed here, but to save space only results on the parameters (ν, ϕ, $1/g$) that are not identifiable in the linear DSGE model are reported. The grid search iterations from Tables B.10 and B.11 show numerically that ν, ϕ and $1/g$ become globally identifiable when the DSGE model is solved with second-order approximations, confirming thus the conjecture from An and Schorfheide [2007]. The reason why in this particular model, quadratic approximations can identify some structural parameters that are not identifiable under the log-linearized model is that ν and ϕ become linearly independent in the nonlinear version of the Philips curve, while g *does* appear in the nonlinear equilibrium equations (government spending).

2.4 Forecasting with Observationally Equivalent Parameter Points

In this section, based on observationally equivalent parameter points obtained from the previous grid search results, we perform several forecasting exercises. Given that in the nonlinear model, all parameters are globally identified, we focus here on the linear model only, where there is evidence for global identi-

fication failure. The results are reported for observationally equivalent points resulted from optimizing the log-likelihood.[14] The algorithm that generates point and density forecasts from a DSGE model is reproduced in Appendix B5.[15]

2.4.1 Correct Specification

In the first forecasting exercise, the forecasts are constructed using indistinguishable parameter values obtained from a correctly specified DSGE model. In this case, data is generated from a model where the monetary policy is given by an output growth specification and estimation is performed using the same model structure.

Table 2.2 reproduces the two observationally equivalent points that are employed in the forecasting exercise. They are selected from the grid search results of Table B.5 given in the appendix and they have clearly distinct values for the unidentified parameters: the inverse elasticity of demand (ν), the price stickiness (ϕ) and the steady state of government spending and ($1/g$). The system matrices A and B from the transition equation are then evaluated using the structural parameter values with ID 6 and 7 from Table 2.2 and the obtained values are given in Appendix B3.[16] Using these values, one step ahead point forecast for Output, Inflation and Interest Rate are constructed

[14]Results obtained from the log-posterior optimization are similar and are available on request.

[15]The algorithm that is applied in the forecasting exercise of Section 2.4 is however slightly different from the standard Bayesian prediction algorithm reproduced in Appendix B5, because the grid search results from the *log-likelihood* optimization are used in prediction. Instead of having a sequence of $\theta^{(i)}$, i= 1,...,n_{sim} posterior draws, we only have one log-likelihood estimate for θ and thus, only point forecasts are computed.

[16]The matrix C, is also used in forecasting, however as it is specified when relating the state variables to the set of observables in the measurement equation, it is always the same and it is not reproduced here.

and the results reported in Table 2.3, below. The three series are measured in percentages.

The results indicate that under correct specification, two distinct parameter points that imply the same value of the log-likelihood, -208.73 (and thus give the same dynamics of the observables), also imply the same system matrices (see matrices $A_{1,1}$, $B_{1,1}$, $A_{1,2}$ and $B_{1,2}$ in Appendix B3) in the state space representation which results in the same point forecast values (column two and three in Table 2.3). The exercise is repeated with a randomly chosen structural parameter with values for ν, ϕ and $1/g$ belonging to an economically admissible range, see column four in Table 2.2, below. This point, being chosen randomly, does not imply a value of -208.73 for the log-likelihood and the system matrices are clearly distinct (see matrices $A_{1,3}$ and $B_{1,3}$ in Appendix B3) from the system matrices obtained previously for the points with ID 6 and 7. As a result, the forecasts implied by the randomly chosen parameter value are different (see column four in Table 2.3) from the previous two sets of forecasts.

Table 2.2: Observationally Equivalent Points - Correct Specification

Parameter	**ID 6**	**ID 7**	**Random**
τ	2	2	2
ν	0.1413	0.1159	0.1200
ϕ	42.8819	53.7948	72.00
$1/g$	0.8575	0.8378	0.9100
ψ_1	1.5	1.5	1.5
ψ_2	1	1	1
ρ_r	0.6	0.6	0.6
ρ_g	0.95	0.95	0.95
ρ_z	0.65	0.65	0.65
r^A	0.4	0.4	0.4
π^A	4	4	4
γ^Q	0.5	0.5	0.5
σ_r	0.002	0.002	0.002
σ_g	0.008	0.008	0.008
σ_z	0.0045	0.0045	0.0045
LogLike	-208.73	-208.73	-

NOTE: The table reports the values of the observationally equivalent points, obtained under correct specification, used in the forecasting exercise. They are the points with ID numbers: 6 and 7 from Table B.5. The last column gives the values of a random point that is *not* observationally equivalent to the points with ID: 6 and 7.

Table 2.3: One Step Ahead Point Forecasts - Correct Specification

Forecast	**Output (6)**	**Output (7)**	**Output (R)**
1	-0.500580	-0.500580	-0.524319
2	-0.506683	-0.506683	-0.530433
3	-0.51139	-0.511392	-0.535149311
4	-0.51515	-0.515155	-0.538917
5	-0.518269	-0.518269	-0.542034
6	-0.520927	-0.520927	-0.544694
7	-0.523257	-0.523257	-0.547025
8	-0.525344	-0.525344	-0.549113
9	-0.527244	-0.527244	-0.551013
10	-0.528995	-0.528995	-0.552764
11	-0.530623	-0.530623	-0.554393
12	-0.532148	-0.532148	-0.555918
13	-0.533582	-0.533582	-0.557352
14	-0.534934	-0.534934	-0.558704
15	-0.536213	-0.536213	-0.559983
	Inflation (6)	**Inflation (7)**	**Inflation (R)**
1	2.843552	2.843542	3.110114
2	2.842874	2.842864	3.109580
3	2.842436	2.842426	3.109234
4	2.842152	2.842142	3.109009
5	2.841967	2.841958	3.108864
6	2.841848	2.841836	3.108769
7	2.841760	2.841760	3.108708
8	2.841719	2.841710	3.108668
9	2.841687	2.841677	3.108642
10	2.841665	2.841656	3.108625
11	2.841651	2.841642	3.108614
12	2.841642	2.841633	3.108607
13	2.841636	2.841627	3.108603
14	2.84163	2.841623	3.108603
15	2.841630	2.841620	3.108598

NOTE: The table reports 15 one step ahead point forecasts for the observables: Output, Inflation and Interest Rate using observationally equivalent parameter values (ID 6 and 7 from Table 2.2) obtained under correct specification of the DSGE model (columns two and three) as well as forecasts obtained using a random (ID R from Table 2.2) parameter value (column four).

Table 2.3: CONTINUED: One Step Ahead Point Forecasts - Correct Specification

Forecast	Interest Rate (6)	Interest Rate (7)	Interest Rate (R)
1	6.103907	6.103899	6.225861
2	6.094307	6.094299	6.216329
3	6.088062	6.088054	6.210127
4	6.084000	6.083992	6.206094
5	6.081359	6.081351	6.203471
6	6.079642	6.079634	6.201767
7	6.078526	6.078518	6.200658
8	6.077801	6.077793	6.199938
9	6.077329	6.077321	6.199470
10	6.077023	6.077015	6.199165
11	6.076823	6.076816	6.198967
12	6.076694	6.076686	6.198839
13	6.076610	6.076602	6.198755
14	6.076555	6.076547	6.198701
15	6.076519	6.076512	6.198665

NOTE: The table reports 15 one step ahead point forecasts for the observables: Output, Inflation and Interest Rate using observationally equivalent parameter values (ID 6 and 7 from Table 2.2) obtained under correct specification of the DSGE model (columns two and three) as well as forecasts obtained using a random (ID R from Table 2.2) parameter value (column four).

2.4.2 Misspecification

The section is motivated by the fact that all DSGE models are approximations, thus they are misspecified in various aspects. The question that is addressed here is whether model misspecification aggravates the consequences of identification failure.

Artificial data is generated according to a true DGP given by an output gap specification, in the monetary policy rule, however estimation is performed under two different misspecification cases: first, assuming an output growth

specification and then an expected inflation specification.

The selected observationally equivalent parameter points used in forecasting are reproduced in Table 2.4 (misspecification with output growth) and Table 2.5 (misspecification with expected inflation). As previously, a random structural parameter is also added to the forecasting exercise, for both misspecification cases.

The results, indicate that even under misspecification, the system matrices implied by observationally equivalent points are the same (see matrices $A_{2,1}$, $B_{2,1}$, $A_{2,2}$, and $B_{2,2}$ in Appendix B3 for the misspecification with output growth case, and matrices $A_{3,1}$, $B_{3,1}$, $A_{3,2}$, and $B_{3,2}$ in Appendix B3 for the misspecification with expected inflation case). As a consequence, the forecasts resulted from these system matrices are also the same (see columns two and three in Tables 2.6 and 2.7.

Table 2.4: Observationally Equivalent Points - Misspecification, Output Growth

Parameter	**ID 12**	**ID 15**	**Random**
τ	2	2	2
ν	0.2505	0.3356	0.1200
ϕ	79.4404	52.5846	33.3300
$1/g$	0.8378	0.9181	0.8000
ψ_1	1.5	1.5	1.5
ψ_2	1	1	1
ρ_r	0.6	0.6	0.6
ρ_g	0.95	0.95	0.95
ρ_z	0.65	0.65	0.65
r^A	0.4	0.4	0.4
π^A	4	4	4
γ^Q	0.5	0.5	0.5
σ_r	0.002	0.002	0.002
σ_g	0.008	0.008	0.008
σ_z	0.0045	0.0045	0.0045
LogLike	-388.93	-388.93	-

NOTE: The table reports the values of the observationally equivalent points, obtained under misspecification, used in the forecasting exercise. They are the points with ID numbers: 12 and 15 from Table B.8. The last column gives the values of a random point that is *not* observationally equivalent to the points with ID: 12 and 15.

Table 2.5: Observationally Equivalent Points - Misspecification, Inflation

Parameter	**ID 6**	**ID 7**	**Random**
τ	2	2	2
ν	0.1060	0.2036	0.1700
ϕ	71.5111	33.1646	35.3300
$1/g$	0.8272	0.8196	0.7200
ψ_1	1.5	1.5	1.5
ψ_2	1	1	1
ρ_r	0.6	0.6	0.6
ρ_g	0.95	0.95	0.95
ρ_z	0.65	0.65	0.65
r^A	0.4	0.4	0.4
π^A	4	4	4
γ^Q	0.5	0.5	0.5
σ_r	0.002	0.002	0.002
σ_g	0.008	0.008	0.008
σ_z	0.0045	0.0045	0.0045
LogLike	-208.28	-208.28	-

NOTE: The table reports the values of the observationally equivalent points, obtained under misspecification, used in the forecasting exercise. They are the points with ID numbers: 6 and 7 from Table B.9. The last column gives the values of a random point that is *not* observationally equivalent to the points with ID: 6 and 7.

Table 2.6: One Step Ahead Point Forecasts - Misspecification, Output Growth

Forecast	**Output (12)**	**Output (15)**	**Output (R)**
1	0.384477	0.384484	0.368643
2	0.372387	0.372393	0.359393
3	0.362007	0.362012	0.351191
4	0.352926	0.352931	0.343776
5	0.344849	0.344852	0.336976
6	0.337560	0.337563	0.330674
7	0.330905	0.330908	0.324789
8	0.324772	0.324774	0.319265
9	0.319075	0.319078	0.314059
10	0.313755	0.313757	0.309142
11	0.308764	0.308766	0.304489
12	0.304066	0.304068	0.300081
13	0.299633	0.299635	0.295900
14	0.295443	0.295444	0.291933
15	0.291476	0.291478	0.288168
	Inflation (12)	**Inflation (15)**	**Inflation (R)**
1	4.855805	4.855720	5.161374
2	4.907939	4.90785	5.212190
3	4.955592	4.955506	5.259014
4	4.999525	4.999440	5.302564
5	5.040315	5.040229	5.343334
6	5.078395	5.078309	5.381676
7	5.114100	5.114013	5.417851
8	5.147689	5.147602	5.452053
9	5.179367	5.179279	5.484441
10	5.209299	5.209210	5.515140
11	5.237621	5.237533	5.544260
12	5.264449	5.264361	5.571895
13	5.289882	5.289793	5.598130
14	5.314005	5.313915	5.623041
15	5.336896	5.336806	5.646698

NOTE: The table reports 15 one step ahead point forecasts for the observables: Output, Inflation and Interest Rate using observationally equivalent parameter values (ID 12 and 15 from Table 2.4) obtained under correct specification of the DSGE model (columns two and three) as well as forecasts obtained using a random (ID R from Table 2.4) parameter value (column four).

Table 2.6: CONTINUED: One Step Ahead Point Forecasts - Misspecification, Output Growth

Forecast	Interest Rate (12)	Interest Rate (15)	Interest Rate (R)
1	5.863216	5.863127	6.226126
2	5.886511	5.886416	6.266017
3	5.912472	5.912371	6.304923
4	5.939879	5.939774	6.342532
5	5.967873	5.967765	6.378680
6	5.995856	5.995744	6.413291
7	6.023421	6.023307	6.446346
8	6.050299	6.050182	6.477862
9	6.076319	6.076200	6.507876
10	6.101377	6.101256	6.536436
11	6.125420	6.125298	6.563600
12	6.148426	6.148302	6.589425
13	6.170396	6.170271	6.613972
14	6.191348	6.191222	6.637300
15	6.211308	6.211180	6.659467

NOTE: The table reports 15 one step ahead point forecasts for the observables: Output, Inflation and Interest Rate using observationally equivalent parameter values (ID 12 and 15 from Table 2.4) obtained under correct specification of the DSGE model (columns two and three) as well as forecasts obtained using a random (ID R from Table 2.4) parameter value (column four).

Table 2.7: One Step Ahead Point Forecasts - Misspecification, Inflation

Forecast	Output (6)	Output (7)	Output (R)
1	1.972836	1.972836	1.972818
2	1.940253	1.940253	1.940243
3	1.908767	1.908767	1.908761
4	1.878505	1.878505	1.878501
5	1.849526	1.849526	1.849523
6	1.821846	1.821846	1.821845
7	1.795453	1.795453	1.795452
8	1.770316	1.770316	1.770315
9	1.746394	1.746394	1.746394
10	1.723641	1.723641	1.723641
11	1.702009	1.702009	1.702009
12	1.681447	1.681447	1.681447
13	1.661905	1.661905	1.661905
14	1.643336	1.643336	1.643336
15	1.625692	1.625692	1.625692
	Inflation (6)	**Inflation (7)**	**Inflation (R)**
1	5.020973	5.021015	5.171180
2	5.021377	5.021419	5.171625
3	5.021632	5.021674	5.171905
4	5.021794	5.021836	5.172084
5	5.021898	5.021940	5.172200
6	5.021965	5.022007	5.172274
7	5.022008	5.022051	5.172322
8	5.022036	5.022079	5.172353
9	5.022055	5.022097	5.172373
10	5.022067	5.022109	5.172386
11	5.022074	5.022116	5.172395
12	5.022079	5.022121	5.172401
13	5.022083	5.022125	5.172404
14	5.022085	5.022127	5.172406
15	5.022086	5.022128	5.172408

NOTE: The table reports 15 one step ahead point forecasts for the observables: Output, Inflation and Interest Rate using observationally equivalent parameter values (ID 6 and 7 from Table 2.5) obtained under correct specification of the DSGE model (columns two and three) as well as forecasts obtained using a random (ID R from Table 2.5) parameter value (column four).

Table 2.7: CONTINUED: One Step Ahead Point Forecasts - Misspecification, Inflation

Forecast	Interest Rate (6)	Interest Rate (7)	Interest Rate (R)
1	6.073802	6.073820	6.136157
2	6.078522	6.078540	6.140937
3	6.081612	6.081629	6.144062
4	6.083628	6.083646	6.146101
5	6.084943	6.084960	6.147430
6	6.085798	6.085816	6.148294
7	6.086355	6.086373	6.148857
8	6.086717	6.086735	6.149223
9	6.086953	6.086971	6.149461
10	6.087106	6.087124	6.149615
11	6.087206	6.087223	6.149716
12	6.087270	6.087288	6.149781
13	6.087312	6.087330	6.149824
14	6.087340	6.087358	6.149851
15	6.087358	6.087375	6.149869

NOTE: The table reports 15 one step ahead point forecasts for the observables: Output, Inflation and Interest Rate using observationally equivalent parameter values (ID 6 and 7 from Table 2.5) obtained under correct specification of the DSGE model (columns two and three) as well as forecasts obtained using a random (ID R from Table 2.5) parameter value (column four).

2.4.3 Observational Equivalent Structures

Section 2.4.1 and Section 2.4.2, treated the case when parameter identification is conditional on the model structure. In these settings, the question was whether there exist a different parameter value within the same DSGE structure that leads to the same dynamics of the observables. In this section, we analyze the identification problem permitting different structures. Here, the question is whether different DSGE structures, for example models with different policy rules, different types of frictions or shock processes, can generate

the same dynamic properties.

Going back to the grid search results, Tables B.5 and B.9 show that different structures can be similar quantitatively. In particular, a DSGE model structure where the monetary policy follows an interest rate rule that reacts to the output gap (i.e. $r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1\pi_t + (1-\rho_r)\psi_2(x_t - g_t) + \epsilon_{rt}$), and a DSGE model structure where that central banks responds to expected inflation rather than current inflation (i.e. $r_t = \rho_r r_{t-1} + (1-\rho_r)\psi_1 E_t\pi_{t+1} + (1-\rho_r)\psi_2(x_t - g_t) + \epsilon_{rt}$), imply the same log-likelihood value of (approximately) -208, leading to (near) observationally equivalence. Thus, in a frequentist approach, it would be hard for the researcher to decide for example between the parameter estimates ID 6 resulted from the output gap specification, or ID 6 resulted from the expected inflation specification. The two points are given below in Table 2.8 and have clearly different values for ν, ϕ and $1/g$.

In contrast to observationally equivalent points belonging to the same model structure, these two points imply distinct system matrices due two different monetary policy rules used in the DSGE model specification (see matrices $A_{1,1}$, $B_{1,1}$, $A_{3,1}$ and $B_{3,1}$ in Appendix B3). As a result, the forecasts resulted from these two structural parameter values are also different, see again the results in column two in Table 2.3 and 2.7. In the output gap specification, the forecasted quarter-to-quarter GDP growth rates are around -0.5%, the annualized quarter-to-quarter inflation rates are around 2.8% and the annualized nominal interest rates are around 6.1% - 6.07%. In contrast, the forecasts obtained from the expected inflation specification vary in the range of 1.9% - 1.6% for the quarter-to-quarter GDP growth rates, the annualized quarter-to-quarter inflation rates are around 5% and the annualized nominal interest rate are around 6.07% - 6.08%.

Table 2.8: Observationally Equivalent Points - Different Structures

Parameter	**ID 6 (O. Gap)**	**ID 6 (Inflation)**
τ	2	2
ν	0.1413	0.1060
ϕ	42.8819	71.5111
$1/g$	0.8575	0.8272
ψ_1	1.5	1.5
ψ_2	1	1
ρ_r	0.6	0.6
ρ_g	0.95	0.95
ρ_z	0.65	0.65
r^A	0.4	0.4
π^A	4	4
γ^Q	0.5	0.5
σ_r	0.002	0.002
σ_g	0.008	0.008
σ_z	0.0045	0.0045
LogLike	-208.73	-208.28

NOTE: The table reports the values of two observationally equivalent points, from two different model structures: one with an output gap specification and another with expected inflation specification. They are the points with ID: 6 from Table B.5 and ID 6: from Table B.9.

2.5 Conclusion

The paper shows that when observational equivalent points belong to the same model structure, global identification failure does not affect the forecasts constructed using dynamic stochastic general equilibrium models - either correctly specified or misspecified. This is a favorable result for policy institutions when the purpose is to construct predictions from DSGE models (where the $ABCD$ matrices are not identifiable locally or globally) that might not be globally identified.

One the other hand, when the observationally equivalent parameter points belong to different DSGE model structures, they can lead to distinct predictions. This result has been illustrated through a small scale DSGE model where two statistically indistinguishable structural parameter estimates, imply distinct forecasts for macroeconomic variables such as: output, inflation and interest rate. One of the parameter estimates resulted from a structure where the monetary policy follows an interest rate rule that reacts to the output gap, and the other, from a structure where the central bank responds to expected inflation rather than current inflation. In fact, in this example, the expected inflation rule is found to be observationally equivalent to a current inflation rule.

In the considered example, failure of global identification within a structure occurs concomitantly with (global) identification failure across structures. One could have examples where the existence of observationally equivalent structures is not necessarily coupled with global identification failure within a structure. Consider for example two model structures that yield the same likelihood, but where both model structures are each globally identified at

the parameter values that optimize the likelihood. Overall, when constructing forecasts based on DSGE models, the results of the paper point to the necessity of taking into account different possible policy scenarios that could be consistent with the data, i.e. observationally equivalent structures.

Chapter 3

Identification Robust Predictive Ability Testing

3.1 Introduction

The paper considers the predictive ability evaluation of models affected by identification deficiencies. Comparing the predictive ability of models where identification might fail, has not yet been examined in the literature, even though many models from empirical predictive evaluation studies can suffer from identification problems. Consider for example the Smooth Transition Autoregressive (STAR) models or the simple ARMA(1,1) model with close to equal autoregresive (AR) and moving-average (MA) parameters. A range of other examples of models that satisfy our identification setup is given in Section 3.2.2.

The class of models we are concerned with, are models where the parameter of interest θ, is of the form $\theta = (\beta, \zeta, \pi)$, where π is identified if and only if $\beta \neq 0$. The parameter ζ is not related to the identification of π, while

$\psi = (\beta, \zeta)$ is always identified. In our setup, the potential source of identification deficiency is an intrinsic characteristic of the model. For example, in the STAR models, one can lose identification in part of the parameter space, because the parameter that measures the slope of the transition and the parameter that gives the location of the transition[1] are not identified when the coefficient on the nonlinear regressor is zero.

Given the structure of the class of models under consideration, when $\beta = 0$, a criterion function[2], used to estimate the parameters of the model, does not depend on π. When β is close to 0, the criterion function is flat with respect to π. As a result, the second derivative matrix of the criterion function is singular or near singular causing the failure of standard asymptotic approximations, as they involve the inverse of this matrix.

The magnitude of $\|\beta\|$ determines the identification strength of the model (or equivalently the parameter π)[3], which as discussed below can be: non-identification, weak identification, semi-strong identification[4] and strong identification. In an in-sample estimation framework, Andrews and Cheng [2012], Andrews and Cheng [2013] and Andrews and Cheng [2014], hereafter AC2012, AC2013 and AC2014, respectively, show that in this class of models, estimators can be inconsistent or consistent with different rates of convergence and have different asymptotic distributions, depending on the identification strength. The asymptotic null distribution of test statistics such as t, QLR

[1] Note that β, π and ζ can be scalars or vectors.

[2] This criterion function can be: maximum likelihood criterion function, least-square criterion function, GMM criterion function, minimum distance criterion function and others.

[3] Throughout the paper, when β is a vector, $\|\cdot\|$ denotes the Euclidean norm, when β is scalar, $\|\cdot\|$ is replaced by the absolute value function.

[4] Semi-strong identification is a technical construct, used to bridge the gap between weak and strong identification. See Section 3.2.1 for details.

or $Wald$ [5] is shown to be nonstandard when the model is non-identified or weakly identified. Once the parameter π is at least semi-strongly identified, the asymptotic null distribution of these tests becomes standard.

In this paper, we analyze the effect of identification loss in an out-of-sample predictive ability evaluation framework. We show that in the considered class of models, under weak and non-identification, the finite-sample distribution of some of the estimators are far from the normal distribution. In the example we consider, that uses a STAR model, these finite sample distributions can be strongly bimodal or uniform. Due to these nonstandard distributions, the forecast errors generated from a model that is not identified are larger than the errors that would have been obtained if the model was (strongly) identified. This affects the null distribution of out-of-sample predictive ability tests, which are not well approximated by the standard normal distribution when one (or both) model(s) considered for predictive ability comparison are not (strongly) identified, in situations in which the parameter estimation error is negligible.

In a numerical example, we show that tests and confidence intervals that employ the standard normal critical value have size distortions and can lead to incorrect inference. The weaker is the identification strength, the lower are the coverage probabilities of confidence intervals (CI) based on the standard normal critical value.

We propose methods to make the out-of-sample predictive ability tests and CIs robust to identification deficiencies. These methods use a different critical value than the standard one and include: a least-favorable critical value and a data dependent critical value. The least favorable critical value

[5] The restrictions tested by these tests, involve in general both ψ and π.

is based on the least favorable possible distribution of the test statistic. For the data dependent critical value, we use an identification category selection procedure to determine the identification strength of π and adjust the critical value accordingly.

In settings where the parameter error is non-negligible, we show that the asymptotic distribution of the predictive ability test in West [1996] is standard even when we allow for π to be only semi-strongly identified.[6] Semi-strong identification does not affect the asymptotic distribution of the test statistic, even though it is known that under semi-strong identification the estimator of π has a slower rate of convergence than the standard (square-root of the sample size) rate.

The paper contributes to the strand of literature that focuses on predictive ability evaluation of forecasting models. Predictive accuracy evaluation topics already considered in the literature include among others: accounting for nested models (Clark and McCracken [2001, 2005], McCracken [2007]), finite-sample predictive ability (Giacomini and White [2006], Clark and McCracken [2009]), multiple model comparison (White [2000], Hansen [2005], Corradi and Distaso [2011], evaluation of conditional quantile forecasts (Giacomini and Komunjer [2005]), tests for density forecasts (Corradi and Swanson [2006], Corradi and Swanson [2006b], Amisano and Giacomini [2007]), predictive ability evaluation in unstable environments (Giacomini and Rossi [2010], Rossi and Sekhposyan [2013]), evaluation of factor-augmented models (Goncalves et al. [2015], Fosten [2016]), allowing for latent target variables (Li and Patton [2013]). For surveys on recent developments in predictive ac-

[6]Semi-strong identification is not covered by the results in West [1996], only the usual strong identification.

curacy methodology see for example Corradi and Swanson [2013], Clark and McCracken [2013], Diebold [2015].

The focus of this paper is on predictive evaluation of models where the assumption of (strong) identification fails. Recent papers such as AC2012, AC2013, AC2014 and Cheng [2015] have analyzed the type of models considered in this paper, in an in-sample estimation framework. AC2012 provide results for general extremum estimators. Under high-level assumptions regarding the behavior of the estimator criterion function, they establish consistency/lack-of-consistency of estimators and determine the asymptotic distributions of estimators under a range of drifting sequences of true distributions. The properties of tests and confidence sets for parameters are analyzed under different identification categories. AC2013 and AC2014 provide a set of primitive conditions for the high level assumptions in AC2012. AC2013 focus on models estimated by log-likelihood criterion functions, while the focus of AC2014 is on moment condition models estimated by GMM. The focus of Cheng [2015] is on *mixed identification*. While AC2012, AC2013 and AC2014 treat a broader class of models with a single source of non-identification, Cheng [2015] provides results for nonlinear regression models with multiple nonlinear regressors and thus multiple sources of non/weak identification.

Related to the problem analyzed in the above mentioned papers is the setting considered in Shi and Phillips [2012], where they allow for the possibility of weak identification in models characterized by a nonlinear cointegrating relationship. Weak identification in their model arises from the presence of a loading coefficient associated to the nonlinear function that might be close to zero. Andrews, Cheng, and Guggenberger [2011] provide generic results used to convert asymptotic results under drifting sequences of parameters into

results that hold uniformly over the parameter space. These results are useful in establishing the asymptotic size in a uniform sense of confidence sets and tests, and they can be applied, among others, to the class of models considered in this papers.

The paper is also broadly related to many other papers on *weak and non-identification* as well as papers on *robust inference in weakly identified models*. Consider for example the literature on robust inference with weakly identified nuisance parameters (see for e.g. Chaudhuri and Zivot [2011], Dufour and Taamouti [2005, 2007], Guggenberger et al. [2012]), or papers on weak identification in macroeconomic models (see for e.g. Andrews and Mikusheva [2015, 2016], Guerron-Quintana et al. [2013], Qu [2014]).

The paper is organized as follows. Section 3.2 introduces the class of models that we focus on, describes the identification problem and outlines the predictive ability evaluation framework. Section 3.3 provides numerical evidence on the implications of identification loss. Section 3.4 establishes asymptotic results under semi-strong identification. Section 3.5 proposes methods to construct robust tests and confidence intervals and applies them in a numerical example. Section 2.5 concludes. The appendix provides proofs and additional numerical results.

3.2 Setup

This section presents the class of models that we consider in our predictive evaluation framework and its identification characteristics. We further give a couple of examples of models used in empirical studies that satisfy our identification setup and outline the predictive ability evaluation framework.

3.2.1 Class of Models and Identification Categories

Consider the class of regression models with additive nonlinearity:

$$y_t = Z_t'\zeta + f(X_t, \pi)'\beta + u_t \tag{3.1}$$

where $y_t \in \mathbb{R}$, $X_t \in \mathbb{R}^{d_X}$, $Z_t \in \mathbb{R}^{d_Z}$ are observed variables, and $u_t \in \mathbb{R}$ is an unobserved error term. The models belonging to (3.1) are nonlinear parametric regressions[7], where $f(X_t, \pi)$ is a smooth nonlinear function with unknown transformation parameter $\pi \in \mathbb{R}^{d_\pi}$. The function $f(X_t, \pi)$ is *known* up to the finite dimensional parameter π. The loading coefficient $\beta \in \mathbb{R}^{d_\beta}$ measures the importance of the nonlinear component, while $\zeta \in \mathbb{R}^{d_\zeta}$ are the coefficients of some linear regressors in the model.

The inherent characteristic of the class of models in (3.1) is that π is identified if and only if $\beta \neq 0$. In fact, $\|\beta\|$ determines the identification strength of π. The parameter β can converge to 0 at various rates or, on the contrary, can be bounded away from 0. The parameter ζ associated to the linear part of the regression is not related to the identification of π, while $\psi = (\beta, \zeta) \in \mathbb{R}^{d_\psi}$ is always identified.[8] The parameters of interest in the model can be collected in a vector $\theta = (\beta, \zeta, \pi) \in \mathbb{R}^{d_\theta}$.

Models of the type given in (3.1) are strongly identified at some points of the parameter space, but might be unidentified or weakly identified at other points of the parameter space. This feature, requires non-standard inference techniques for this class of models. Moreover, the large sample properties

[7]The ARMA(1,1) model with equal autoregressive and moving-average coefficients present similar identification issues as (3.1), but it is linear.

[8]This setup also allows for cases where a model is reparametrized to convert it into the framework considered here.

of estimators, tests and confidence sets differ according to the identification properties of the model (see for e.g. AC2012, Cheng [2015]).

When $\beta = 0$, the criterion function used to estimate the parameters of the model, does not depend on π, while when β is close to zero, the criterion function is flat with respect to π. This means that the second derivative matrix of the criterion function is singular or near singular, causing the failure of standard asymptotic approximations, as they involve the inverse of this matrix.

In order to model the different identification strengths of π, alternative asymptotic approximations along drifting sequences of true parameters need to be considered. To this end, suppose that the true value of the parameter is: $\theta_T = (\beta_T, \zeta_T, \pi_T)$, where $T \geq 1$ indexes the sample size. As shown in AC2012 the behaviour of the extremum estimators and test statistics associated to models in (3.1), depends on the magnitude of $\|\beta_T\|$ and varies across three categories of sequences $\{\beta_T : T \geq 1\}$, which, following the terminology in AC2012, are:

Category I (a): when $\beta_T = 0 \; \forall \; T \geq 1$

Category I (b): when $\beta_T \neq 0$ and $T^{1/2}\beta_T \to b \in \mathbb{R}^{d_\beta}$

Category II: when $\beta_T \to 0$ and $T^{1/2}\|\beta_T\| \to \infty$

Category III: when $\beta_T \to \beta_0 \neq 0$, $\beta_0 \in \mathbb{R}^{d_\beta}$

When the sequences $\{\beta_T\}$ satisfy the conditions from category I (a), category I (b), category II and category III, the parameter π is, respectively, unidentified, weakly identified, semi-strongly identified and strongly identified. Note that when π is weakly identified, β_T converges to zero at the same rate as $1/\sqrt{T}$,

while when π is semi-strongly identified, β_T converges to zero slower than $1/\sqrt{T}$.

As pointed out in AC2012, $\theta = (\beta, \zeta, \pi)$ does not need to completely determine the distribution of the data. Thus, an additional parameter ϕ can be introduced, which is an infinite dimensional nuisance parameter, such that $\gamma = (\theta, \phi) \in \Gamma$ fully determines the distribution of the data.[9] Therefore, in addition to the drifting sequences $\{\beta_T : T \geq 1\}$, we can allow other parameters to change with the sample size and obtain uniform results not only over β but also over γ.

Based on the identification categories described above, define the following sequences of true parameters $\{\gamma_T\}$, where $\{\gamma_T\} = \{(\theta_T, \phi_T)\}$:

1. $\Gamma(\gamma_0) = \{\{\gamma_T \in \Gamma : T \geq 1\} : \gamma_T \to \gamma_0 \in \Gamma\}$
2. $\Gamma(\gamma_0, 0, b) = \{\{\gamma_T\} \in \Gamma(\gamma_0) : \beta_0 = 0 \text{ and } T^{1/2}\beta_T \to b \in (\mathbb{R} \cup \pm\infty)\}$
3. $\Gamma(\gamma_0, \infty, \omega_0) = \{\{\gamma_T\} \in \Gamma(\gamma_0) : T^{1/2}\beta_T \to \infty \text{ and } \beta_T/\|\beta_T\| \to \omega_0 \in \mathbb{R}^{d_\beta}\}$

where $\gamma_0 = (\beta_0, \zeta_0, \pi_0, \phi_0) \in \Gamma$ is the limit of the sequences $\{\gamma_T\}$. In the notation above, the 0 in $\Gamma(\gamma_0, 0, b)$ means $\beta_0 = 0$ while the ∞ in $\Gamma(\gamma_0, \infty, \omega_0)$ stands for $T^{1/2}\|\beta_0\| \to \infty$.

The sequences $\{\gamma_T\}$ in the set $\Gamma(\gamma_0, 0, b)$, are sequences for which $\{\beta_T\}$ is close to 0 and are in identification category I and II. The sequences $\{\gamma_T\}$ in the set $\Gamma(\gamma_0, \infty, \omega_0)$ are sequences for which $\{\beta_T\}$ is more distant from 0, and are in identification category II and III. Note that both sets $\Gamma(\gamma_0, 0, b)$ and $\Gamma(\gamma_0, \infty, \omega_0)$ contain sequences $\{\gamma_T\}$ from identification category II. What

[9]For example, in nonlinear regression models estimated by least squares, θ indexes a finite-dimensional feature of the distribution of the errors, (such as its variance), while ϕ indexes the remaining characteristics of the distribution of the errors, which may be infinite dimensional.

distinguishes these two sets is whether $T^{1/2}\|\beta_T\| \to \|b\|$ with $\|b\| < \infty$ or $\|b\| = \infty$.[10]

3.2.2 Examples

The class of models considered in the paper includes many well known models which appear frequently in empirical studies.

Example 1. *Smooth Transition Autoregressive (STAR) Models*

The model can be written as follows:

$$\begin{gathered} y_t = X_t'\zeta + X_t'\beta \cdot f(V_t, \pi) + u_t, \\ \text{with } X_t = (1, y_{t-1}, \ldots, y_{t-p})' \text{ and } V_t = y_{t-d} \end{gathered} \tag{3.2}$$

The function $f(V_t, \pi)$ is known and the values of the parameters p and d are also assumed to be known, with $1 \leq d \leq p$. The literature usually considers two different forms for the transition function $f(V_t, \pi) \in [0, 1]$: the logistic function

$$f(V_t, \pi) = (1 + \exp\left[-\pi_1^*(V_t - \pi_2^*)\right])^{-1} \tag{3.3}$$

or the exponential function

$$f(V_t, \pi) = 1 - \exp\left[-\pi_1^*(V_t - \pi_2^*)^2\right] \tag{3.4}$$

where $\pi = (\pi_1^*, \pi_2^*)' \in \mathbb{R}^2$. The parameter $\pi_1^* > 0$ measures the slope of the transition, while the parameter π_2^* the location of the transition. When β=0, the parameter π is not identified. The model can be estimated using an

[10] Note that b indexes the magnitude of β.

estimator criterion function such as the least square or maximum likelihood criterion function. The first derivative of these criterion functions with respect to π are proportional to β, when β is close to zero. Hence, the criterion functions are flat in the direction of π. For applications of this model see for e.g. Luukkonen et al. [1988], Terasvirta and Anderson [1992], Teräsvirta [1994], Ferrara, Marcellino, and Mogliani [2015], among others.

Example 2. *Nonlinear Regression with Endogeneity*

Consider the following nonlinear regression model with endogenous regressors:

$$y_t = \beta \cdot f(X_{1t}, \pi) + X_{2t}'\zeta + u_t \text{ with instruments } Z_t \text{ independent of } u_t \quad (3.5)$$

where X_{1t} is a nonlinear regressor, X_{2t} is a linear regressor, and Z_t are the instrumental variables which are assumed to be strong. Identification failure in this model can arise due to the nonlinearity introduced by the function f. This model can be estimated using a GMM criterion function. In this example, the endogeneity of the regressors is not crucial for our identification problem. The reason to consider it here is to be able to give example of different extremum estimators that can be used. An application of this model is considered in Areosa et al. [2011], where they model inflation rate in Brazil and find strong support for a nonlinear specification of the Philips curve. In their empirical study, the nonlinear function f is a logistic function.

Example 3. *ARMA(1,1) model*

It is a well known result that common autoregressive (AR) and moving average (MA) roots lead to identification failure in the ARMA(1,1) model. In other words, the AR and MA parameters are not identified when they are

equal. This occurs when the series is white noise. Consider the following ARMA(1,1) model:

$$y_t = (\pi + \beta)y_{t-1} + \epsilon_t - \pi\epsilon_{t-1} \tag{3.6}$$

when $\beta = 0$ the model becomes $y_t = \pi y_{t-1} + \epsilon_t - \pi\epsilon_{t-1}$ which is equivalent to $y_t = \epsilon_t$. Thus, when $\beta = 0$ the AR parameter $\pi + \beta$ and the MA parameter π are not identified. Ansley and Newbold [1980] and Nelson and Startz [2007] show in simulations that AR and MA parameters that are close in value, lead to bias, variance and size problems. The ARMA(1,1) example fits our identification setup, even though there is no nonlinear function involved.

Other examples of models that satisfy our identification setup include: mixed data sampling (MIDAS) regressions in empirical finance (see e.g. Ghysels, Sinko, and Valkanov [2007]), autoregressive distributed lag models, continuous transition structural change models, continuous transition threshold autoregressive models (see e.g. Chan and Tsay [1998]), models with correlated random coefficients (see e.g. Andrews [2001]), nonlinear binary choice models, probit models with endogeneity and possibly weak instruments (see e.g. Rivers and Vuong [1988]).[11] The focus of this paper is on the subset of models that satisfy the identification setup described above and are also used in prediction.

Our framework does not cover all models that have identification deficiencies at some points in the parameter space. For example, models such as regime switching models, abrupt transition structural change models, and abrupt transition threshold autoregressive models, such as in Hansen [2000], Liu and Shao [2003], Elliott and Müller [2007, 2014], Qu and Perron [2007] are

[11] For other examples and references for models from this class, see AC2012.

not considered here and are left for future research.

3.2.3 Predictive Evaluation Framework

We are interested in comparing the predictive ability of two models where one or both belong to the class described in Section 3.2.1, in situations where one or both models are affected by identification deficiencies. Here, we outline the predictive evaluation framework with notations corresponding to the case when both models belong to (3.1).

Consider thus the forecasting model 1:

$$y_{t+h} = Z_{1t}'\zeta_1 + f_1(X_{1t}, \pi_1)'\beta_1 + u_{t+h} \tag{3.7}$$

and a benchmark model, model 2:

$$y_{t+h} = Z_{2t}'\zeta_2 + f_2(X_{2t}, \pi_2)'\beta_2 + \epsilon_{t+h} \tag{3.8}$$

where $y_t \in \mathbb{R}$, $X_{1t} \in \mathbb{R}^{d_{X_1}}$, $Z_{1t} \in \mathbb{R}^{d_{Z_1}}$, $X_{2t} \in \mathbb{R}^{d_{X_2}}$, $Z_{2t} \in \mathbb{R}^{d_{Z_2}}$ are observed variables, $u_t \in \mathbb{R}$ and $\epsilon_t \in \mathbb{R}$ are unobserved error terms and h is the forecast horizon.

Alternatively, the benchmark forecast could be also considered a non-model based forecast such as those obtained from different surveys. In this paper, we maintain a non-nested framework, i.e. we consider the case where $W_{2t} = (X_{2t}, Z_{2t})$ is not nested within $W_{1t} = (X_{1t}, Z_{1t})$.

In order to test the predictive accuracy of the model in (3.7) relative to the accuracy of the model in (3.8), we can formulate the null hypothesis of

equal unconditional predictive ability as:

$$H_0 : E[g(u_{t+h}) - g(\epsilon_{t+h})] = 0 \tag{3.9}$$

where u_{t+h} and ϵ_{t+h} are the forecast errors of model (3.7) and model (3.8) and $g(\cdot)$ is some differentiable (or non-differentiable) loss function. This null hypothesis tests the equality of expected forecast error losses resulted from the two competing models. The alternative can be two sided (the two models do not perform equally well) or one-sided (one of the models performs better).

The forecast errors of model (3.7) and model (3.8) can be obtained by using a rolling, recursive or a fixed window estimation scheme. The notations in this paper correspond to the rolling estimation scheme, but recursive and fixed estimation cases could be also considered. Thus, the sample of size T is divided into an in-sample part R and an out-of-sample part P and the models' parameters are re-estimated for each rolling window of length R, indexed by t with $R \leq t \leq T$.

Suppose the parameters of interest of the model in (3.7), $\theta_1 = (\zeta_1, \beta_1, \pi_1)$ are estimated by minimizing a criterion function $Q_R(\theta_1)$, that depends on the observables and the sample size R, over a parameter space $\Theta_1 \subseteq \mathbb{R}^{d_{\theta_1}}$. The criterion function is minimized successively over rolling windows of data. Analogously, θ_2 is estimated by minimizing a criterion function $Q_R(\theta_2)$ over an optimization parameter space $\Theta_2 \subseteq \mathbb{R}^{d_{\theta_2}}$.

Given the parameter estimates, the forecast errors based on a rolling window of observations from model (3.7) and model (3.8) are given by:

$$\hat{u}_{t+h} = y_{t+h} - Z'_{1t}\hat{\zeta}_{1,R,t} - f(X_{1t}, \hat{\pi}_{1,R,t})'\hat{\beta}_{1,R,t} \quad \text{where } R \leq t \leq T \tag{3.10}$$

and

$$\hat{\epsilon}_{t+h} = y_{t+h} - Z_{2t}'\hat{\zeta}_{2,R,t} - f(X_{2t}, \hat{\pi}_{2,R,t})'\hat{\beta}_{2,R,t} \quad \text{where } R \leq t \leq T \tag{3.11}$$

Thus, an out-of-sample predictive ability test statistic such as the Diebold-Mariano-West (DMW) statistic can be formulated as:

$$S_P = \frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} (g(\hat{u}_{t+h}) - g(\hat{\epsilon}_{t+h})) \tag{3.12}$$

and used to test the null given in (3.9).

The main difference of our setup to the classical DMW framework, is the dependence of the test statistic on the estimates of the parameters π_1 and π_2, where π_1 and/or π_2 might not be strongly identified.

3.3 Numerical Evidence: Implications of Identification Loss

In this section we illustrate numerically the impact of identification issues that arise naturally in the class of models described in Section 3.2.1, on the finite-sample densities of estimators, as well as on the finite-sample densities of out-of sample tests and critical values.

3.3.1 Effect on Estimators

As an example of a model from the class (3.1), consider the following LSTAR(1):

$$y_t = \zeta_1 + \zeta_2 y_{t-1} + \beta \cdot f(y_{t-1}, \pi) + u_t \tag{3.13}$$

where $f(x, \pi) = x(1 + \exp[-10(x - \pi)])^{-1}$ and $u_t \sim N(0, 1)$.

In this simulation exercise, for illustration purposes, we use the constant 10 for the parameter that measures the slope of the transition. The identification issues are already evident when only one parameter (the location of the transition) appears in the logistic function. Estimating a second parameter in the logistic function will only aggravate the identification problems.

Figure 3.1 to 3.4 provide the finite sample densities of the maximum likelihood estimators for ζ_1, ζ_2, β, and π, respectively, when the true value of π, $\pi_0 = -1.5$. The number of simulation repetitions is 500, and the sample size is $T = 500$. Each subplot gives the densities of estimators for $b = [0, 2, 4, 10]$ where $b = \sqrt{T}\beta$ indexes the magnitude of β. The values of $b = [0, 2, 4, 10]$ correspond to $\beta = [0, 0.09, 0.18, 0.45]$. The true values of ζ_1 and ζ_2 are set to -1 and 0.5 respectively.

Figures 3.1 and 3.2 show that, the finite-sample densities of the estimators of ζ_1 and ζ_2 are not far from a normal distribution, for all values of b. In contrast, the results for the estimators of β and π are very different, the finite-sample distributions being very far from a normal distribution for smaller values of b.

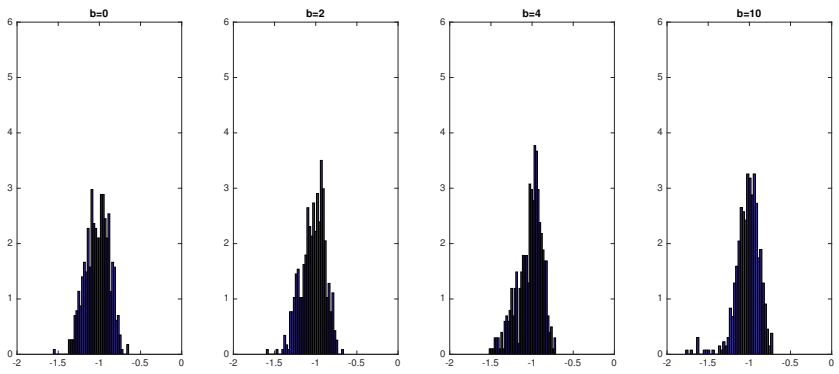

Figure 3.1: Finite sample densities of the estimator of ζ_1 when $\pi_0 = -1.5$.

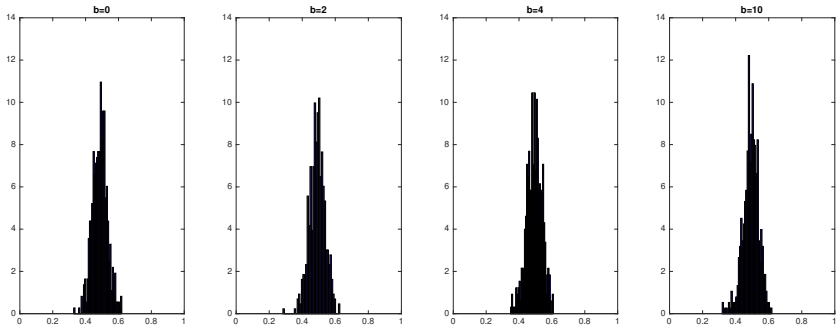

Figure 3.2: Finite sample densities of the estimator of ζ_2 when $\pi_0 = -1.5$.

Figure 3.3 shows that the maximum likelihood estimator of β has a clear bimodal distribution when $b = 0$ (non-identification of π), $b = 2$ and $b = 4$ (weak identification of π). It starts to resemble a normal distribution centered around the true value $\beta_0 = 0.45$ once b is increased to 10. Regarding the distribution of the estimator of π, Figure 3.4 indicates that it is almost uniformly distributed on the interval [-4, -0.5], for a true value pf $\pi_0 = -1.5$ when $b = 0$. When $b = 2$ it still resembles a uniform distribution, with a very slight buildup of mass that starts forming around the true value. The buildup of mass becomes more evident when $b = 4$, and the distribution of the maximum likelihood estimator of π is not far from a normal distribution centered around the true value once $b = 10$[12].

Similar results on the densities of estimators when $\pi_0 = -3$ are reported in Figures C.1 to C.4, in the Appendix.

[12]The asymptotic and finite sample results in AC2012, AC2013 and AC2014 confirm the results obtained in this section.

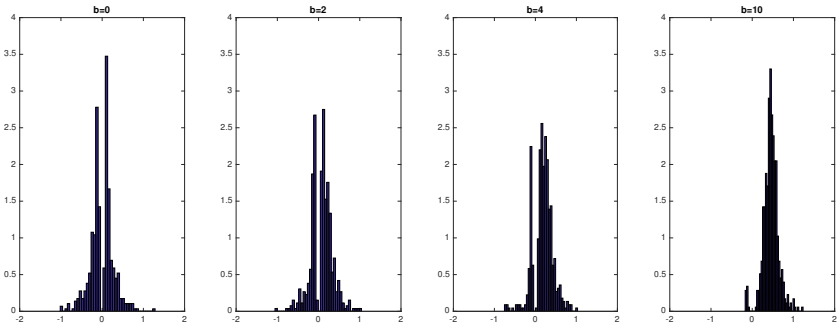

Figure 3.3: Finite sample densities of the estimator of β when $\pi_0 = -1.5$.

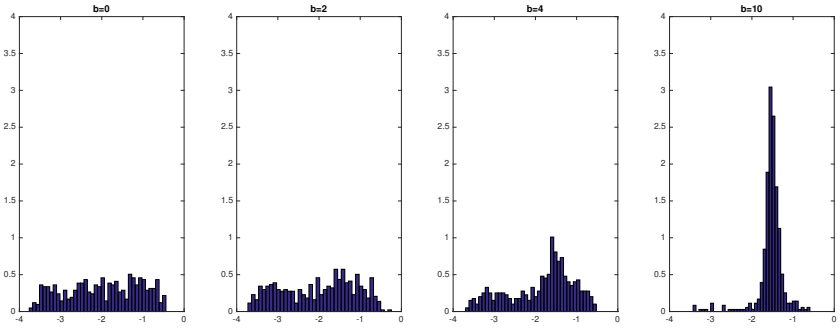

Figure 3.4: Finite sample densities of the estimator of π when $\pi_0 = -1.5$.

3.3.2 Effect on Out-of-Sample Tests

This section focuses on the case in which the parameter estimation error is negligible. In this simulation design, we generate data according to the following two data generating processes (DGPs):

$$y_t = \beta_1 y_{t-1}(1 + \exp[-c_1(y_{t-1} - \pi_1)])^{-1} + v_{1t} \tag{3.14}$$

and

$$y_t = \beta_2 y_{t-1}(1 + \exp[-c_2(y_{t-1} - \pi_2)])^{-1} + v_{2t} \tag{3.15}$$

For the DGP in 3.14, we successively consider the following values for b_1, the parameter that indexes the magnitude of β_1, $b_1 = [4, 6, 8, 10, 12]$. In finite-samples, $b_1 = \beta_1\sqrt{R}$. Thus, with an in-sample size of $R = 500$[13], the b_1 values correspond to true values of $\beta_{0,1} = [0.18, 0.26, 0.35, 0.45, 0.53]$. In the DGP in 3.15, b_2 is fixed to 12 which implies that $\beta_{0,2} = 0.53$, a value for which, according to the simulation results of the previous section, the model is already strongly identified. The true values of the other parameters are set to $c_{0,1} = c_{0,2} = 10$ and $\pi_{0,1} = \pi_{0,2} = -1.5$ and the error terms v_{1t} and v_{2t} are drawn from a $N(0, 1)$ distribution. To simplify computations, in 3.14 and 3.15 we only consider the nonlinear part of the LSTAR(1) model.

Note that in this simulation exercise, we are not interested in results given by very small b_1 values, such as $b_1 = 0$ or $b_1 = 2$, because in that case β_1 is small too, and one would not be interested in forecastig with such small values of β_1.

The two competing models considered for forecast evaluation are the model in 3.14, which can have different strengths of identification depending on b_1, and the model in 3.15, which is always identified. When comparing the predictive ability of 3.14 and 3.15, based for example on the mean square forecast errors (MSE), the null of equal predictive ability should be satisfied. The number of simulation repetitions in the computation of the out-of-sample predictive ability statistic is 500.

Figure 3.5 provides the finite-sample densities of the out-of-sample test statistic, under the null of equal predictive ability, for different b_1 values in 3.14. For smaller b_1 values, the model 1 (equation 3.14) is not (strongly) identified

[13]In this example, the in-sample size R is kept at the same value of 500 as the size of the full sample in the previous subsection. The out-of-sample size is $P = 150$.

and the parameters are not correctly estimated. Thus, the forecast errors generated from this model are larger than the errors obtained from model 2 (equation 3.15). As a consequence, the null distribution of the out-of-sample predictive evaluation test is biased to the right and it is not well approximated by the standard normal distribution. A standard normal critical value could lead to misleading inference. For example, when $b_1 = 4$, the 0.95 quantile of the distribution of the test statistic is 3.11, whereas for $b_1 = 12$, the 0.95 quantile is roughly 2.

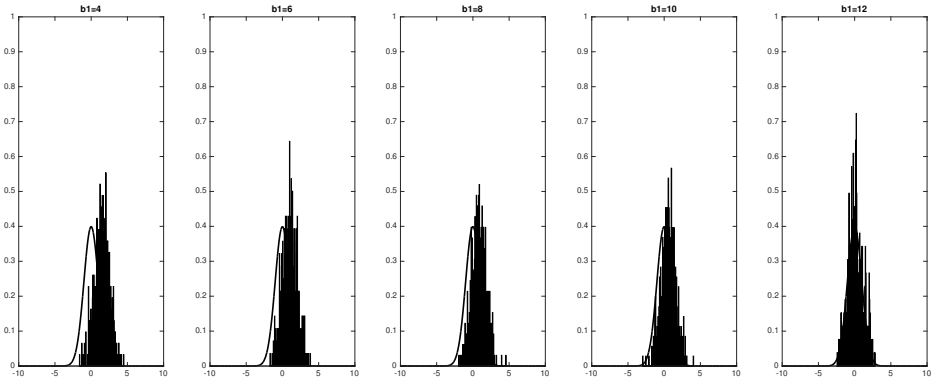

Figure 3.5: Finite sample densities of the Diebold-Mariano test statistic

Figure 3.6 reports the finite sample coverage probabilities of the nominal 95% standard confidence intervals, constructed based on the standard normal critical value, for different b_1 values for model 1[14]. The confidence intervals are constructed by inverting the test statistic. For a number of 500 simulation repetitions, we collect all values of the statistic that in absolute value are less than 1.96, for each discrete value of b_1. Then, the coverage probability is obtained as the proportion of instances in which the value of the statistic is contained in the standard confidence interval. The smaller is the b_1 value (i.e.

[14]The discrete values of b_1 for which computations are made run from 4 to 12 with a grid of 0.2.

the lower is the identification strength), the lower is the coverage probability of the standard confidence interval. The smallest finite-sample coverage probabilities are around 0.81 for $b_1 = 4$. As b_1 increases, the coverage probabilities are progressively approaching 0.95.

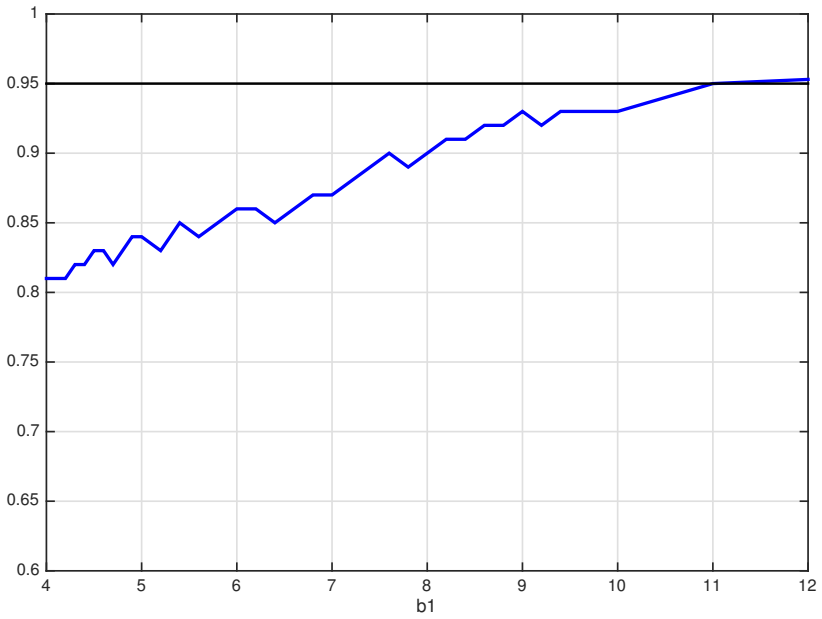

Figure 3.6: Coverage Probabilities of Standard CIs

3.4 Asymptotic Results under Semi-Strong Identification

We are interested in the asymptotic distribution of the statistic S_P, when one or both models belong to the class in 3.1, and thus can suffer from identification deficiencies.

When π_1 and π_2 are non-identified or weakly identified, it is known from the in-sample asymptotic results of AC2012 that their extemum estimators are inconsistent. In consequence, the framework in West [1996] cannot be applied in order to derive the asymptotic distribution of S_P.[15] Finite-sample distributions of out-of-sample tests, can still be analysed in this case through simulation, as in Section 3.3.[16] When π_1 and π_2 are strongly identified, the asymptotic distribution of S_P is of course standard.

In this section, we derive the asymptotic distribution of S_P when π_1 and π_2 are semi-strongly identified[17] and show that the asymptotic distribution of the out-of-sample test under the null of equal predictive ability is standard, even though we allow π_1 and π_2 to be only semi-strongly identified. This is different from West [1996] where all parameters are assumed to be strongly identified. Semi-strong identification does not affect the asymptotic distribution of S_P, even though in this case, $\hat{\pi}_{1R}$ and $\hat{\pi}_{2R}$ have a slower rate of convergence, $\sqrt{R}\|\beta_{1R}\| \ll \sqrt{R}$ and $\sqrt{R}\|\beta_{2R}\| \ll \sqrt{R}$, than it is standard.[18] This is in conformity with the asymptotic results in AC2012 and Cheng [2015], regarding the asymptotic distributions of in-sample statistics such as *t, Wald, QLR*, which are shown to have a standard distribution under the null, once π is at least in the semi-strong identification category. These statistics are shown to have non-standard distributions, only when π is non-identified or weakly identified.

We derive our results for the case when both models belong to the class

[15] More precisely, Theorem 4.1 in West [1996] cannot be applied.

[16] Note that a STAR model with π non-identified or weakly identified with very small beta estimates is not of interest in a prediction evaluation framework. This is not however the case for the ARMA(1,1) model.

[17] When π is semi-strongly identified, its extremum estimator is consistent, see Theorem 3.2 in Andrews and Cheng [2012].

[18] See Theorem 3.2 in Andrews and Cheng [2012].

of models under consideration and both are semi-strongly identified. When only one of the competing models belongs to this class, the same result obtains and the derivations corresponding to the standard model (the one without identification deficiencies) are as in West [1996].

The following assumptions are required to derive our result:

Assumption 1 (Parameter space):

(i) The true parameter space corresponding to model 1, denoted Θ_1^*, lies in the interior of the optimization parameter space Θ_1.

(ii) The true parameter space corresponding to model 2, denoted Θ_2^*, lies in the interior of the optimization parameter space Θ_2.

Assumption 1 ensures that the true value of the parameter cannot lie on the boundary of the optimization parameter space. This way, we exclude possible boundary effects and the focus is solely on the effects of identification deficiencies.

Assumption 2 (Criterion function):

(i) If $\beta_1 = 0$, $Q_R(\theta_1)$ does not depend on π_1, $\forall\ \theta_1 = (\beta_1, \zeta_1, \pi_1) = (0, \zeta_1, \pi_1) \in \Theta_1$, $\forall\ R \geq 1$.

(ii) If $\beta_2 = 0$, $Q_R(\theta_2)$ does not depend on π_2, $\forall\ \theta_2 = (\beta_2, \zeta_2, \pi_2) = (0, \zeta_2, \pi_2) \in \Theta_2$, $\forall\ R \geq 1$.

Assumption 2 is a key assumption in the paper. It is the main characteristic of the class of models under consideration. Under this assumption, the criterion function is flat with respect to π_1 (or π_2) when β_1 (or β_2) is close to zero and the second derivative matrix of the criterion function is singular or near singular.

The next three assumptions, are similar to assumptions D1, D2, and D3 in AC2012 and concern the quadratic expansion that the criterion function

must satisfy when π_1 and π_2 are semi-strongly identified, and, the behavior of the first and second derivatives of the criterion function.

Assumption 3 (Quadratic expansion of the criterion function):
(i) The sample criterion function $Q_R(\theta_1)$ has a quadratic expansion in θ_1 around the true value θ_{1R}, i.e.

$$\begin{aligned} Q_R(\theta_1) = & Q_R(\theta_{1R}) + DQ_R(\theta_{1R})'(\theta_1 - \theta_{1R}) \\ & + \frac{1}{2}(\theta_1 - \theta_{1R})' D^2 Q_R(\theta_{1R})(\theta_1 - \theta_{1R}) + R_R(\theta_1) \end{aligned} \tag{3.16}$$

where $DQ_R(\theta_{1R}) \in \mathbb{R}^{d_{\theta_1}}$ is a stochastic generalized first derivative vector, $D^2 Q_R(\theta_{2R}) \in \mathbb{R}^{d_{\theta_1} \times d_{\theta_1}}$ is a generalized second partial derivative matrix and $R_R(\theta_1)$ is a remainder term.
(ii) The sample criterion function $Q_R(\theta_2)$ has a quadratic expansion in θ_2 around the true value θ_{2R}, i.e.

$$\begin{aligned} Q_R(\theta_2) = & Q_R(\theta_{2R}) + DQ_R(\theta_{2R})'(\theta_2 - \theta_{2R}) \\ & + \frac{1}{2}(\theta_2 - \theta_{2R})' D^2 Q_R(\theta_{2R})(\theta_2 - \theta_{2R}) + R_R(\theta_2) \end{aligned} \tag{3.17}$$

where $DQ_R(\theta_{2R}) \in \mathbb{R}^{d_{\theta_2}}$ is a stochastic generalized first derivative vector, $D^2 Q_R(\theta_{2R}) \in \mathbb{R}^{d_{\theta_2} \times d_{\theta_2}}$ is a generalized second partial derivative matrix and $R_R(\theta_2)$ is a remainder term.

The next assumption requires good behavior of the rescaled generalized second derivative of $Q_R(\theta_1)$ and $Q_R(\theta_2)$, to eliminate its singularity when $\beta_{1R} \to 0$ and $\beta_{2R} \to 0$, which occurs when π_1 and π_2 are semi-strongly identified.

Assumption 4 (Behavior of the generalized second derivative):
(i) $J_{1R} = B^{-1}(\beta_{1R}) D^2 Q_R(\theta_{1R}) B^{-1}(\beta_{1R}) \xrightarrow{p} J_1(\gamma_{0,1}) \in \mathbb{R}^{d_{\theta_1} \times d_{\theta_1}}$, where $J_1(\gamma_{0,1})$

is nonsingular and symmetric.

(ii) $J_{2R} = B^{-1}(\beta_{2R})D^2Q_R(\theta_{2R})B^{-1}(\beta_{2R}) \xrightarrow{p} J_2(\gamma_{0,2}) \in \mathbb{R}^{d_{\theta_2} \times d_{\theta_2}}$, where $J_2(\gamma_{0,2})$ is nonsingular and symmetric.

The matrices $B(\beta_1)$ and $B(\beta_2)$, defined in the appendix, are used to normalize $D^2Q_R(\theta_{1R})$ and $D^2Q_R(\theta_{2R})$ so that they are nonsingular asymptotically.

The following assumption requires the rescaled generalized first derivative to satisfy a CLT.

Assumption 5 (Behavior of the generalized first derivative):

(i) $R^{1/2}B^{-1}(\beta_{1R})DQ_R(\theta_{1R}) \xrightarrow{d} G^*(\gamma_{0,1}) \sim N(0_{d_{\theta_1}}, V(\gamma_{0,1}))$ for a $d_{\theta_1} \times d_{\theta_1}$ symmetric positive definite matrix $V(\gamma_{0,1})$.

(ii) $R^{1/2}B^{-1}(\beta_{2R})DQ_R(\theta_{2R}) \xrightarrow{d} G^*(\gamma_{0,2}) \sim N(0_{d_{\theta_2}}, V(\gamma_{0,2}))$ for a $d_{\theta_2} \times d_{\theta_2}$ symmetric positive definite matrix $V(\gamma_{0,2})$.

The next two assumptions are standard as in West [1996]. They require measurability and differentiability of the loss function g, impose conditions on the moments of g and impose stationarity and mixing conditions.

Assumption 6 (i) In some neighborhood N_1 around θ_1, with probability one, $g(u_{t+h})$ is measurable and twice continuously differentiable with respect to θ_1.

(ii) In some neighborhood N_2 around θ_2, with probability one, $g(\epsilon_{t+h})$ is measurable and twice continuously differentiable with respect to θ_2.

(iii) For all t, there exists a constant $D < \infty$, such that $\sup_{\theta_1 \in N_1} \|\nabla^2_{\theta_1} g(u_{t+h})\| \leq n_{1t}$, for a measurable function n_{1t} for which $E(n_{1t}) < D$.

(iv) For all t, there exists a constant $D < \infty$, such that $\sup_{\theta_2 \in N_2} \|\nabla^2_{\theta_2} g(\epsilon_{t+h})\| \leq n_{2t}$, for a measurable function n_{2t} for which $E(n_{2t}) < D$.

Assumption 7 (i) For some $d > 1$, $\sup_t E\|[\nabla_{\theta_1} g(u_{t+h}),$

$g(u_{t+h}), \nabla_{\theta_2} g(\epsilon_{t+h}),$

$g(\epsilon_{t+h}),]' \|^{4d} < \infty$, where $\| \cdot \|$ is the Euclidean norm.

(ii)$[(\nabla_{\theta_1} g(u_{t+h}) - E(\nabla_{\theta_1} g(u_{t+h})), (g(u_{t+h}) - E(g(u_{t+h}))),$

$(\nabla_{\theta_2} g(\epsilon_{t+h}) - E(\nabla_{\theta_2} g(\epsilon_{t+h})), (g(\epsilon_{t+h}) - E(g(\epsilon_{t+h}))),]'$ is strong mixing with mixing coefficients of size $-3d/(d-1)$.

(iii)$[\nabla_{\theta_1} g(u_{t+h}), g(u_{t+h}), \nabla_{\theta_2} g(\epsilon_{t+h}), g(\epsilon_{t+h})]'$ is covariance stationary.

(iv) Let $\Gamma_{t+h} = g(u_{t+h}) - g(\epsilon_{t+h})$ and let $V_\epsilon = \sum_{j=-\infty}^{+\infty} E(\Gamma_{t+h} - E(\Gamma_{t+h}))(\Gamma_{t+h-j} - E(\Gamma_{t+h-j}))$, then V_ϵ is positive definite.

The last assumption is standard in the predictive ability testing literature and concerns the relative rate of increase of T, R and P.

Assumption 8 $P, R \to \infty$ as $T \to \infty$ and $\lim_{T\to\infty} \frac{P}{R} = \tilde{\pi}$ with $0 \leq \tilde{\pi} \leq \infty$.

The following result provides the asymptotic distribution of S_P for the case in which π_1 and π_2 are semi-strongly identified. The proof of the theorem is given in Appendix 1.

Theorem 1. *Let Assumptions 1-8 hold. Under* H_0,

$$S_P \xrightarrow{d} N(0, \Omega)$$

where

$$\Omega = V_\epsilon + \lambda_1 D'_{\theta_1} V_{W_1} D_{\theta_2} + \lambda_1 D'_{\theta_1} V_{W_2} D_{\theta_2} + 2\lambda_2 D_{\theta_1} C_{\epsilon W_1} - 2\lambda_2 D_{\theta_2} C_{\epsilon W_2} - 2\lambda_1 D'_{\theta_1} C_{W_1 W_2} D_{\theta_2}$$

with:

$$V_\epsilon = \sum_{j=-\infty}^{+\infty} E[(g(u_{t+h}) - g(\epsilon_{t+h})) - E(g(u_{t+h}) - g(\epsilon_{t+h}))$$
$$\times (g(u_{t+h+j}) - g(\epsilon_{t+h+j})) - E(g(u_{t+h}) - g(\epsilon_{t+h}))']$$

$$V_{W_1} = -J_1^{-1}(\gamma_{0,1})B^{-2}(\beta_{1R})\Big(\sum_{j=-\infty}^{+\infty} E(m_{1,t}m'_{1,t+j})\Big)B^{-2}(\beta_{1R})(-J_1^{-1}(\gamma_{0,1}))$$

$$V_{W_2} = -J_2^{-1}(\gamma_{0,2})B^{-2}(\beta_{2R})\Big(\sum_{j=-\infty}^{+\infty} E(m_{2,t}m'_{2,t+j})\Big)B^{-2}(\beta_{2R})(-J_2^{-1}(\gamma_{0,2}))$$

$$C_{\epsilon W_1} = \Big(\sum_{j=-\infty}^{+\infty} E[(g(u_{t+h}) - g(\epsilon_{t+h}))m'_{1,t+j}]\Big)B^{-2}(\beta_{1R})(-J_1^{-1}(\gamma_{0,1}))$$

$$C_{\epsilon W_2} = \Big(\sum_{j=-\infty}^{+\infty} E[(g(u_{t+h}) - g(\epsilon_{t+h}))m'_{2,t+j}]\Big)B^{-2}(\beta_{2R})(-J_2^{-1}(\gamma_{0,2}))$$

$$C_{W_1W_2} = -J_1^{-1}(\gamma_{0,1})B^{-2}(\beta_{1R})\Big(\sum_{j=-\infty}^{+\infty} E(m_{1,t}m'_{2,t+j})\Big)B^{-2}(\beta_{2R})(-J_2^{-1}(\gamma_{0,2}))$$

and $D_{\theta_1} = E(\nabla_{\theta_1} g(u_{t+h}))$, $D_{\theta_2} = E(\nabla_{\theta_2} g(\epsilon_{t+h}))$.

The parameters λ_1 and λ_2 are defined as follows. In a rolling estimation

scheme, we have that $\lambda_1 = (1 - \frac{1}{3\tilde{\pi}})$ and $\lambda_2 = (1 - \frac{1}{2\tilde{\pi}})$, in the case where $1 < \tilde{\pi} < \infty$ and, $\lambda_1 = (\tilde{\pi} - \frac{\tilde{\pi}^2}{3})$ and $\lambda_2 = \frac{\tilde{\pi}}{2}$, in the case where $\tilde{\pi} \leq 1$. If we consider a recursive estimation scheme, $\lambda_1 = 2(1 - \frac{1}{\tilde{\pi}}\ln(1 + \tilde{\pi}))$ and $\lambda_2 = 1 - \frac{1}{\tilde{\pi}}\ln(1 + \tilde{\pi})$. For a fixed window estimation scheme, the parameters λ_1 and λ_2 are $\lambda_1 = \tilde{\pi}$, $\lambda_2 = 0$.

Theorem 1 shows that even though π_1 and π_2 might be only semi-strongly identified, the asymptotic distribution of S_P, based on the class of models defined in (3.1), is the same as in West [1996]. Note that the results in West [1996] are derived assuming OLS estimation, while here we assume a maximum likelihood estimation framework.

3.5 Robust Inference

As shown in the numerical finite-sample results (Section 3.3), a standard strong-identification critical value can lead to misleading inference, when π_1 (and/or π_2) are not (strongly) identified. In this section, we construct robust critical values that take into account the possible loss of (strong) identification. A robust critical value in our testing framework is larger than the standard strong identification critical value. Robust confidence intervals can also be constructed by inverting the test statistic using the robust critical value. Here the focus is on cases in which the parameter estimation error is negligible. Robust critical values in non-negligible parameter estimation settings can be obtained by bootstrap inference. This is considered in future research.

3.5.1 Data Dependent Critical Values

The first step in the construction of robust critical values is to employ an identification selection procedure (ICS) that uses the data to determine whether $\|b\| < \infty$ (i.e. non-identification/weak identification) or $\|b\| = \infty$ (i.e. semi-strong/strong identification).[19]

The identification selection procedure distinguishes between weak and semi-strong identification of π based on the following statistic:[20]

$$ICS_R = (R\hat{\beta}_R' \hat{\Sigma}_{\beta\beta,R}^{-1} \hat{\beta}_R / d_\beta)^{1/2} \tag{3.18}$$

where $\hat{\Sigma}_{\beta\beta,R}$ is the the upper left $d_\beta \times d_\beta$ block of $\hat{\Sigma}_R$ - the estimator of the variance-covariance matrix of $\hat{\theta}_R$, and R is the in-sample size[21]. Note that when $\beta_R = O(\frac{1}{\sqrt{R}})$, (i.e. when π is weakly identified), ICS_R is $O_p(1)$.

We select the weak identification category if

$$ICS_R \leq k_R \tag{3.19}$$

and the semi-strong identification category, otherwise.[22] The sequence of constants $k_R : R \geq 1$ are tuning parameters such that:

$$i)\ k_R \to \infty \quad \text{and} \quad ii)\ k_R / R^{1/2} \to 0. \tag{3.20}$$

[19] This identification selection procedure is also employed in Andrews and Cheng [2012] and Cheng [2015] and is related to methods proposed in Andrews [1999] and Andrews and Soares [2010].

[20] For ease of notation, the indices 1 and 2 corresponding to model 1 and 2 are omitted in this subsection.

[21] In order to decrease computation times, when computing the ICS statistic, a fixed estimation scheme is used.

[22] As ICS_R is $O_p(1)$ when π is weakly identified, one consistently selects the weak identification category, provided k_R diverges to infinity.

For example, k_R can be taken $k_R = (\ln R)^{1/2}$, which is analogous to the Bayesian information criterion (BIC) penalty term, to satisfy i) and ii).

To understand why the second condition on k_R, (ii), is also needed, consider a strong identification case when β_R is bounded away from 0, say $\beta_R = 2$. In this case, $\hat{\beta}_R$ converges to 2 in probability and ICS_R diverges to infinity at rate $\sqrt{R}$. To ensure ICS_R is larger than k_R, we need k_R to diverge at a rate slower than $\sqrt{R}$, thus the second condition $k_R/R^{1/2} \to 0$.

Using the identification selection procedure described above, a robust critical value, at nominal level $1-\alpha$, $\hat{c}_{R,1-\alpha}$ is defined as:

$$\hat{c}_{R,1-\alpha} = \begin{cases} c^{LF}_{1-\alpha} & \text{if } ICS_R \leq k_R \\ c^{\infty}_{1-\alpha} & \text{if } ICS_R > k_R \end{cases} \tag{3.21}$$

where $c^{\infty}_{1-\alpha}$ is the standard strong identification critical value, and $c^{LF}_{1-\alpha}$ is the least favorable (LF) critical value, that is large enough for all identification categories:

$$c^{LF}_{1-\alpha} = \max\left\{ \sup_{l \in L} c_{1-\alpha}(l),\ c^{\infty}_{1-\alpha} \right\}. \tag{3.22}$$

In (3.22), $c_{1-\alpha}(l)$ is the $1-\alpha$ quantile of the statistic under weak/non-identification, and $l = (b, \gamma) \in L$, with $L = \left\{ l = (b, \gamma) : \|b\| < \infty, \gamma \in \Gamma \right\}$. Note that one could directly use, $c^{LF}_{1-\alpha}$, the least favorable critical value as the robust critical value, without the identification selection procedure, but the associated confidence interval is typically overly long, and not as informative in the strong identification case. The ICS procedure improves on the LF critical value.

The robust critical value, $\hat{c}_{R,1-\alpha}$, can be further improved by employing null-imposed and/or plug-in versions of it. A null imposed critical value ex-

ploits the knowledge of the null hypothesis value of a restriction, $r(\theta)$, implied by the test, such that the null is $H_0 : r(\theta) = v$ for $v \in r(\Theta)$. For example, in a context where the null hypothesis specifies a value π^* for π, then the supremum in (3.22) needs to be taken only over the l values for which $\pi = \pi^*$. In this case, L is replaced by $L(v) = \{l = (b, \gamma) \in L : \|b\| < \infty,\ r(\theta) = v\}$ in (3.22).[23]

The plug-in version of $\hat{c}_{R,1-\alpha}$ can be employed when part of $\gamma = (\zeta, \beta, \pi, \phi)$ is unknown under H_0, but can be consistently estimated. This critical value replaces elements of γ with consistent estimators in the formula in (3.22), and the supremum over the set L can be reduced to a supremum over $\hat{L}_R$ (or $L(v) \cap \hat{L}_R$), where $\hat{L}_R = \{l = (b, \gamma) \in L : \gamma = (\hat{\zeta}_R, \beta, \pi, \phi)\}$, as ζ is consistently estimated by $\hat{\zeta}_R$.[24]

3.5.2 Simulation: Coverage Probabilities based on Robust Critical Values

Consider again the simulation design from Section 3.3.2. We apply the methods described above to construct robust critical values and thus obtain robust tests. We consider both the least favorable critical value $c^{LF}_{1-\alpha}$ as well as the data dependent critical value $\hat{c}_{R,1-\alpha}$.

The least favorable critical value $c^{LF}_{1-\alpha}$ is obtained as follows. Fix $\pi_{0,1}$ in 3.14 successively to one of the values $\pi_{0,1} = [-3.5, -3, -2.5, -2, -1.5]$[25]. For each of these $\pi_{0,1}$, run the b_1 values, from 4 to 10 with a grid of 0.2, obtain the

[23] The null imposed critical value is not particularly useful in our setup, as the null involves the loss differential, see equation (3.9), and does not impose a particular value for the parameters.

[24] The plug in version of $\hat{c}_{R,1-\alpha}$ is more suitable to our testing framework.

[25] These are values from the true parameter space of π.

distribution of the test statistic and compute the 0.95 quantile. The maximum of these quantiles over all values of b_1 and $\pi_{0,1}$ gives the least favorable critical value. Note that in this simulation exercise, a value of $\zeta = (\zeta_1, \zeta_2)'$ is not required in order to obtain $c_{1-\alpha}^{LF}$. Because ζ can be consistently estimated, in applications where ζ appears, one could plug-in the estimated value of ζ in place of ζ_0.

Figure 3.7 reports the finite-sample coverage probabilities of confidence intervals constructed based on the least favorable critical value as a function of b_1. The least favorable critical value is obtained for $b_1 = 4$ and $\pi_{0,1} = -2$, and has a value of 3.15. The robust confidence interval is obtained by inverting the test statistic using the least favorable critical value. The coverage probabilities are closer to 0.95 for smaller values of b_1. For larger b_1 values, the coverage probabilities are 1, as the robust confidence interval based on the least favorable critical value is overly long when model 1 is strongly identified.

The data dependent robust critical value $c_{1-\alpha}^{LF}$ that employs an identification category selection procedure does not lead to coverage probabilities of 1, once π_1 is at least semi-strongly identified because the robust critical value switches to the standard critical value.

Figure 3.8 reports the coverage probabilities computed based on the data dependent robust critical value. For $R = 500$, $k_R = (\ln R)^{1/2} = 2.49$. This value is compared with the value of the ICS_R statistic obtained over the grid values of b_1. The identification category selection statistic becomes larger than k_R at $b_1 = 11.2$. Using the data dependent robust critical value the coverage probabilities are closer to 0.95 for larger values of b_1, where π_1 is at least semi strongly identified. In this procedure, the transition from the least favorable critical value to the standard critical value is not continuous.

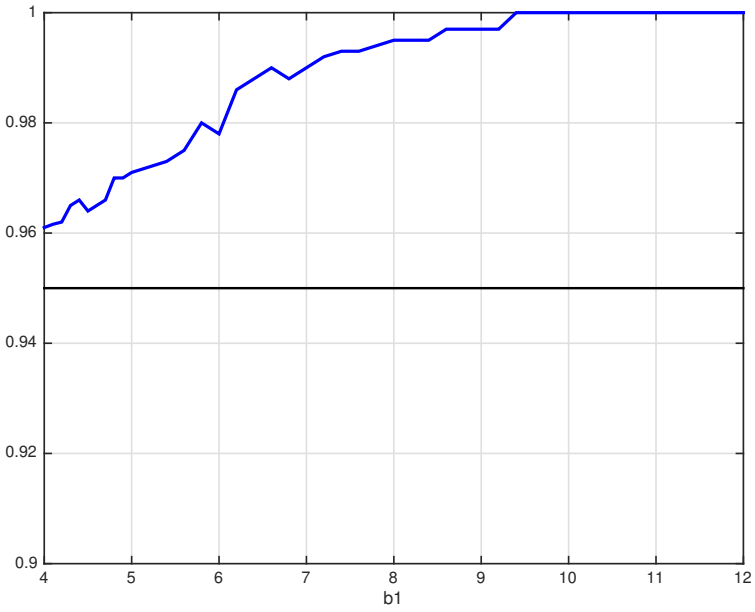

Figure 3.7: Coverage probabilities of robust CIs with least favorable critical value

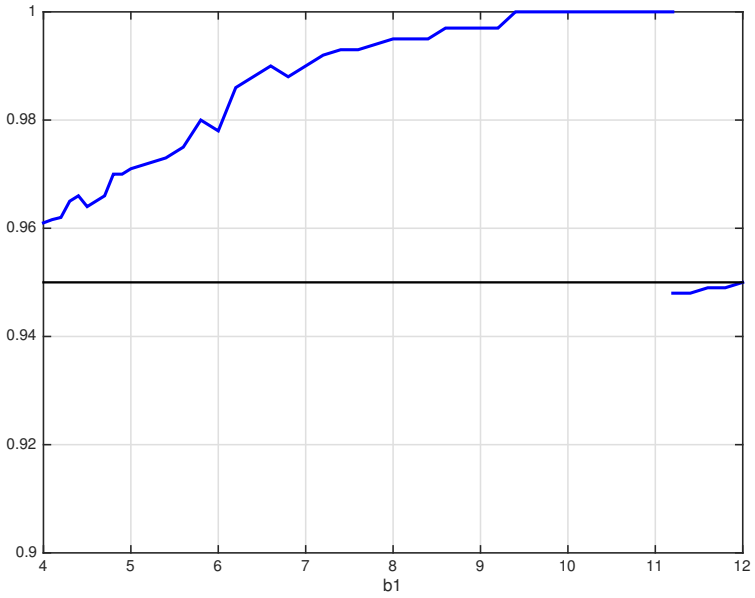

Figure 3.8: Coverage probabilities of robust CIs with data dependent critical values

3.6 Conclusion

The paper analyzes the predictive accuracy evaluation of models affected by identification deficiencies. The focus is on a class of models that are strongly identified in some part of the parameter space but non-identified or weakly identified in another part of the parameter space. The potential source of strong identification loss is an explicit part of the model.

Numerical results show that when identification is lost, the finite-sample distributions of estimators are far from the normal distribution, they can be bimodal or uniform, for example. As a result, the forecast errors obtained under identification loss are larger than the errors obtained under strong identification and, in situations in which the parameter estimation is negligible, the null distribution of out-of-sample predictive ability tests is not well approximated by the standard normal distribution. Thus, out-of-sample tests and confidence intervals obtained by inverting tests that employ the standard normal critical value have size distortions and can lead to misleading inference.

We propose two methods to construct robust critical values for this problem. The first is based on the least favorable possible distribution of the test statistic under the null, whereas the second is data dependent and uses an identification category selection procedure to determine the identification strength.

In settings where the parameter estimation error is non-negligible, the paper shows that the asymptotic distribution of the out-of-sample predictive ability test in West [1996] is standard, even when one allows for the model(s) to be only semi-strongly identified.

Appendix A

Appendix to Chapter One

Table A.1: GMM estimates for α obtained under the true loss function

R, P	$\alpha_0 = \mathbf{0.2}$	$\alpha_0 = \mathbf{0.4}$	$\alpha_0 = \mathbf{0.5}$	$\alpha_0 = \mathbf{0.6}$	$\alpha_0 = \mathbf{0.8}$
R=250, P=150	0.1986	0.3982	0.5012	0.6041	0.8043
R=250, P=200	0.1984	0.4007	0.4989	0.5994	0.8020
R=250, P=250	0.1992	0.3994	0.5017	0.6009	0.8001
R=200, P=250	0.1997	0.3989	0.4985	0.6002	0.8003
R=300, P=200	0.1982	0.4006	0.5000	0.6001	0.8023

NOTE: The table reports the average estimates for the asymmetry parameter α across 1000 Monte Carlo simulations for different values of the true asymmetry parameter α_0. R is the size of the rolling window used to construct the forecasts and P is the size of the evaluation sample.

Table A.2: Rejection frequencies for the J-test when the forecast evaluation is done under the true loss function

R, P	$\alpha_0 = \mathbf{0.2}$	$\alpha_0 = \mathbf{0.4}$	$\alpha_0 = \mathbf{0.5}$	$\alpha_0 = \mathbf{0.6}$	$\alpha_0 = \mathbf{0.8}$
R=250, P=150	0.0460	0.0420	0.0440	0.0520	0.0540
R=250, P=200	0.0430	0.0510	0.0380	0.0390	0.0470
R=250, P=250	0.0320	0.0460	0.0400	0.0300	0.0350
R=200, P=250	0.0260	0.0250	0.0340	0.0380	0.0350
R=300, P=200	0.0350	0.0450	0.0470	0.0340	0.0390

NOTE: The table reports the percentage of rejections of the null of rationality at the 5% nominal level for different values of the true asymmetry parameter α_0. The forecaster's true loss function belongs to $L_1(\epsilon_{t+1}; p, \alpha) = [\alpha + (1 - 2\alpha) \cdot 1(\epsilon_{t+1} \leq 0)] \cdot |\epsilon_{t+1}|^p$. R is the size of the rolling window used to construct the forecasts and P is the size of the evaluation sample.

Table A.3: GMM estimates for α obtained under a misspecified loss function

R, P	a_0=**-5**	a_0=**-3**	a_0=**-1**	a_0=**1**	a_0= **3**	a_0= **5**
R=250, P=150	0.0482	0.1326	0.3461	0.6532	0.8621	0.7769
R=250, P=200	0.0494	0.1331	0.3470	0.6540	0.8606	0.7769
R=250, P=250	0.0509	0.1346	0.3488	0.6516	0.8608	0.7751
R=200, P=250	0.0538	0.1382	0.3498	0.6501	0.8573	0.7778
R=300, P=200	0.0482	0.1321	0.3484	0.6522	0.8647	0.7745

NOTE: The table reports the average estimates for the asymmetry parameter α across 1000 Monte Carlo simulations for different values of the Linex loss asymmetry parameter and for different sizes of the rolling window R and forecast evaluation sample P.

Table A.4: Rejection frequencies for the J-test when the forecast evaluation is done under a misspecified loss function

R, P	a_0=**-5**	a_0=**-3**	a_0=**-1**	a_0=**1**	a_0= **3**	a_0= **5**
R=250, P=150	0.0870	0.0610	0.0360	0.0510	0.2130	0.3950
R=250, P=200	0.0930	0.0580	0.0450	0.0400	0.2950	0.5130
R=250, P=250	0.1010	0.0710	0.0320	0.0310	0.3260	0.5890
R=200, P=250	0.1060	0.0470	0.0300	0.0310	0.2780	0.6030
R=300, P=200	0.0870	0.0670	0.0380	0.0420	0.3070	0.4880

NOTE: The table reports the percentage of rejections of the null of rationality at the 5% nominal level for different values of the Linex loss asymmetry parameter and for different sizes of the rolling window R and forecast evaluation sample P. The forecaster's true loss function is the Linex loss: $L_2(\epsilon_{t+1}; a) = \exp(a \cdot \epsilon_{t+1}) - a \cdot \epsilon_{t+1} - 1$.

Table A.5: Empirical Size and Power for the two tests

		J-**Stat**	M_T-**Stat**
$g(x) = x^2$			
	$\delta = 0.2$	0.186	0.436
	$\delta = 0.5$	0.244	0.662
	$\delta = 1$	0.242	0.592
	$\delta = 2$	0.310	0.800
$g(x) = arctan(x)$			
	$\delta = 0.2$	0.296	0.610
	$\delta = 0.5$	0.302	0.724
	$\delta = 1$	0.298	0.762
	$\delta = 2$	0.320	0.812
$g(x) = exp(x)$			
	$\delta = 0.2$	0.290	0.616
	$\delta = 0.5$	0.324	0.736
	$\delta = 1$	0.362	0.856
	$\delta = 2$	0.430	0.816
	$\delta = 0$	0.104	0.116

NOTE: The table reports test rejection frequencies at a 10% significance level using a sample size of $T = 500$. The number of Monte Carlo replications is $M = 1000$. For the M_T-stat the number of bootstrap replications for each Monte Carlo replication is $B = 100$.

Table A.6: J-test based on Median Forecasts

	Instrument	α	**std.err.**	t**-Stat**	J**-Stat**	**CV at 10%**	p**-value**
Real GNP/GDP							
	Case 1	0.4630	0.0488	-0.7586	1.2149	2.71	0.2704
	Case 2	0.4759	0.0484	-0.4966	0.5967	2.71	0.4398
	Case 3	0.4651	0.0483	-0.7216	0.2062	2.71	0.6498
	Case 4	0.4715	0.0483	-0.5890	0.0640	2.71	0.8003
	Case 5	0.4683	0.0483	-0.6570	1.6501	4.60	0.4382
	Case 6	0.4662	0.0482	-0.7010	2.5526	6.25	0.4659
Price Index GNP/GDP							
	Case 1	0.5541	0.0471	1.1480	0.4478	2.71	0.5034
	Case 2	0.5676	0.0463	1.4601	2.6596	2.71	0.1029
	Case 3	0.5626	0.0460	1.3622	0.9246	2.71	0.3363
	Case 4	0.5610	0.0459	1.3312	0.6171	2.71	0.4321
	Case 5	0.5621	0.0458	1.3549	0.9456	4.60	0.6233
	Case 6	0.5587	0.0457	1.2841	1.9772	6.25	0.5772
Consumption							
	Case 1	0.2760	0.0502	-4.4603	5.9011	2.71	0.0151
	Case 2	0.3075	0.0522	-3.6884	0.7563	2.71	0.3845
	Case 3	0.2815	0.0503	-4.3453	4.1666	2.71	0.0412
	Case 4	0.3057	0.0519	-3.7425	0.3205	2.71	0.5713
	Case 5	0.2732	0.0498	-4.5532	6.0723	4.60	0.0480
	Case 6	0.2651	0.0492	-4.7755	7.1208	6.25	0.0681

NOTE: The table reports the asymmetry parameter α estimates, corresponding standard errors, values of the t-statistic testing H_0: α=0.5, the values of the J statistic and its critical values and p-values for Cases 1-6. Case 1: constant and lagged errors, Case 2: constant and absolute lagged errors Case 3: constant and lagged change in actual values, Case 4: constant and lagged change in forecasts, Case 5: constant and lagged errors plus lagged change in actual values, Case 6: constant and lagged errors and lagged change in actual values and lagged change in forecasts. The sample size is $T = 177$ for Output and Prices and $T = 126$ for Consumption.

Table A.7: M_T-test based on Median Forecasts

	Instrument	M_T**-Stat**	**Boot.CV at** 5%	**Boot.CV at** 10%
Real GNP/GDP				
	Case 1	0.0229	0.0581	0.0466
	Case 2	0.0224	0.0439	0.0439
	Case 3	0.0200	0.0457	0.0391
	Case 4	0.0164	0.0554	0.0493
	Case 5	0.0235	0.0470	0.0370
	Case 6	0.0208	0.0577	0.0494
Price Index GNP/GDP				
	Case 1	0.0316	0.0340	0.0303
	Case 2	0.0321	0.0257	0.0219
	Case 3	0.0322	0.0419	0.0302
	Case 4	0.0336	0.0319	0.0281
	Case 5	0.0308	0.0315	0.0305
	Case 6	0.0313	0.0371	0.0303
Consumption				
	Case 1	0.0152	0.0378	0.0338
	Case 2	0.0114	0.0397	0.0347
	Case 3	0.0171	0.0512	0.0400
	Case 4	0.0123	0.0410	0.0377
	Case 5	0.0220	0.0540	0.0471
	Case 6	0.0239	0.0520	0.0464

NOTE: The table reports the values of the M_T test statistic and its Bootstrap Critical Values at 5% and at 10% for Cases 1-6. Case 1: constant and lagged errors, Case 2: constant and absolute lagged errors Case 3: constant and lagged change in actual values, Case 4: constant and lagged change in forecasts, Case 5: constant and lagged errors plus lagged change in actual values, Case 6: constant and lagged errors and lagged change in actual values and lagged change in forecasts. The sample size is $T = 177$ for Output and Prices and $T = 126$ for Consumption. The block length is 5. The blocks are overlapping with an overlap length of 2. The number of bootstrap replications is $B = 100$.

Table A.8: J-test based on Mean Forecasts

	Instrument	α	**std.err.**	**t-Stat**	J**-Stat**	**CV at 10%**	p**-value**
Real GNP/GDP							
	Case 1	0.4556	0.0488	-0.9100	0.7695	2.71	0.3804
	Case 2	0.4688	0.0484	-0.6450	1.5282	2.71	0.2164
	Case 3	0.4550	0.0481	-0.9348	0.2026	2.71	0.6526
	Case 4	0.4603	0.0479	-0.8283	0.0144	2.71	0.9045
	Case 5	0.4591	0.0479	-0.8532	0.8986	4.60	0.6381
	Case 6	0.4567	0.0478	-0.9055	1.8447	6.25	0.6053
Price Index GNP/GDP							
	Case 1	0.5532	0.0472	1.1252	0.7229	2.71	0.3952
	Case 2	0.5543	0.0468	1.1586	0.3312	2.71	0.5650
	Case 3	0.5619	0.0461	1.3422	1.1376	2.71	0.2862
	Case 4	0.5602	0.0460	1.3082	0.7403	2.71	0.3896
	Case 5	0.5610	0.0460	1.3262	1.1972	4.60	0.5496
	Case 6	0.5578	0.0459	1.2594	2.2457	6.25	0.5230
Consumption							
	Case 1	0.3041	0.0523	-3.7489	5.4346	2.71	0.0197
	Case 2	0.3325	0.0538	-3.1113	0.3351	2.71	0.5627
	Case 3	0.3152	0.0527	-3.5047	2.8750	2.71	0.0900
	Case 4	0.3325	0.0538	-3.1158	0.0037	2.71	0.9515
	Case 5	0.3043	0.0522	-3.7486	5.4401	4.60	0.0659
	Case 6	0.2976	0.0518	-3.9110	6.4760	6.25	0.0906

NOTE: The table reports the asymmetry parameter α estimates, corresponding standard errors, values of the t-statistic testing H_0: α=0.5, the values of the J statistic and its critical values and p-values for Cases 1-6. Case 1: constant and lagged errors, Case 2: constant and absolute lagged errors Case 3: constant and lagged change in actual values, Case 4: constant and lagged change in forecasts, Case 5: constant and lagged errors plus lagged change in actual values, Case 6: constant and lagged errors and lagged change in actual values and lagged change in forecasts. The sample size is $T = 177$ for Output and Prices and $T = 126$ for Consumption.

Table A.9: J-test based on Range Forecasts

	Instrument	α	**std.err.**	**t-Stat**	J**-Stat**	**CV at 10%**	p**-value**
Real GNP/GDP							
	Case 1	0.4606	0.0487	-0.8091	1.2459	2.71	0.2643
	Case 2	0.4749	0.0483	-0.5192	0.7263	2.71	0.3941
	Case 3	0.4651	0.0482	-0.7732	0.2062	2.71	0.6498
	Case 4	0.4628	0.0482	-0.6253	0.2251	2.71	0.6352
	Case 5	0.4664	0.0481	-0.6995	1.6993	4.60	0.4276
	Case 6	0.4643	0.0480	-0.7430	2.6056	6.25	0.4565
Price Index GNP/GDP							
	Case 1	0.5454	0.0470	0.9656	1.1797	2.71	0.2774
	Case 2	0.5492	0.0466	1.0556	1.1358	2.71	0.2865
	Case 3	0.5503	0.0460	1.0922	0.8136	2.71	0.3671
	Case 4	0.5464	0.0460	1.0077	0.2856	2.71	0.5931
	Case 5	0.5486	0.0460	1.0575	1.2741	4.60	0.5289
	Case 6	0.5454	0.0459	0.9894	2.3338	6.25	0.5061
Consumption							
	Case 1	0.2662	0.0494	-4.7295	6.7588	2.71	0.0093
	Case 2	0.3042	0.0520	-3.7668	0.9609	2.71	0.3270
	Case 3	0.2826	0.0503	-4.3194	3.7277	2.71	0.0535
	Case 4	0.3069	0.0520	-3.7135	0.0219	2.71	0.8824
	Case 5	0.2665	0.0493	-4.7306	6.7658	4.60	0.0339
	Case 6	0.2584	0.0487	-4.9597	7.8038	6.25	0.0502

NOTE: The table reports the asymmetry parameter α estimates, corresponding standard errors, values of the t-statistic testing H_0: α=0.5, the values of the J statistic and its critical values and p-values for Cases 1-6. Case 1: constant and lagged errors, Case 2: constant and absolute lagged errors Case 3: constant and lagged change in actual values, Case 4: constant and lagged change in forecasts, Case 5: constant and lagged errors plus lagged change in actual values, Case 6: constant and lagged errors and lagged change in actual values and lagged change in forecasts.The sample size is $T = 177$ for Output and Prices and $T = 126$ for Consumption.

Table A.10: M_T-test based on Mean Forecasts

	Instrument	M_T**-Stat**	**Boot. CV at** 5%	**Boot. CV at** 10%
Real GNP/GDP				
	Case 1	0.0054	0.0460	0.0406
	Case 2	0.0078	0.0612	0.0445
	Case 3	0.0028	0.0578	0.0454
	Case 4	0.0010	0.0453	0.0397
	Case 5	0.0049	0.0600	0.0461
	Case 6	0.0023	0.0542	0.0512
Price Index GNP/GDP				
	Case 1	0.0298	0.0306	0.0228
	Case 2	0.0312	0.0380	0.0292
	Case 3	0.0297	0.0361	0.0266
	Case 4	0.0311	0.0326	0.0284
	Case 5	0.0283	0.0297	0.0260
	Case 6	0.0281	0.0309	0.0263
Consumption				
	Case 1	0.0130	0.0382	0.0350
	Case 2	0.0087	0.0429	0.0365
	Case 3	0.0137	0.0467	0.0387
	Case 4	0.0086	0.0485	0.0376
	Case 5	0.0188	0.0539	0.0436
	Case 6	0.0196	0.0547	0.0481

NOTE: The table reports the values of the M_T test statistic and its Bootstrap Critical Values at 5% and at 10% for Cases 1-6. Case 1: constant and lagged errors, Case 2: constant and absolute lagged errors Case 3: constant and lagged change in actual values, Case 4: constant and lagged change in forecasts, Case 5: constant and lagged errors plus lagged change in actual values, Case 6: constant and lagged errors and lagged change in actual values and lagged change in forecasts. The sample size is $T = 177$ for Output and Prices and $T = 126$ for Consumption. The block length is 5. The blocks are overlapping with an overlap length of 2. The number of bootstrap replications is $B = 100$.

Table A.11: M_T-test based on Range Forecasts

	Instrument	M_T**-Stat**	**Boot. CV at** 5%	**Boot. CV at** 10%
Real GNP/GDP				
	Case 1	0.0114	0.0601	0.0408
	Case 2	0.0113	0.0522	0.0414
	Case 3	0.0082	0.0477	0.0403
	Case 4	0.0047	0.0572	0.0487
	Case 5	0.0116	0.0601	0.0505
	Case 6	0.0084	0.0697	0.0538
Price Index GNP/GDP				
	Case 1	0.0367	0.0337	0.0272
	Case 2	0.0380	0.0377	0.0298
	Case 3	0.0382	0.0357	0.0298
	Case 4	0.0399	0.0416	0.0328
	Case 5	0.0365	0.0348	0.0287
	Case 6	0.0381	0.0417	0.0338
Consumption				
	Case 1	0.0116	0.0431	0.0361
	Case 2	0.0074	0.0360	0.0306
	Case 3	0.0124	0.0477	0.0376
	Case 4	0.0071	0.0418	0.0360
	Case 5	0.0178	0.0500	0.0432
	Case 6	0.0187	0.0530	0.0485

NOTE: The table reports the values of the M_T test statistic and its Bootstrap Critical Values at 5% and at 10% for Cases 1-6. Case 1: constant and lagged errors, Case 2: constant and absolute lagged errors Case 3: constant and lagged change in actual values, Case 4: constant and lagged change in forecasts, Case 5: constant and lagged errors plus lagged change in actual values, Case 6: constant and lagged errors and lagged change in actual values and lagged change in forecasts. The sample size is $T = 177$ for Output and Prices and $T = 126$ for Consumption. The block length is 5. The blocks are overlapping with an overlap length of 2. The number of bootstrap replications is $B = 100$.

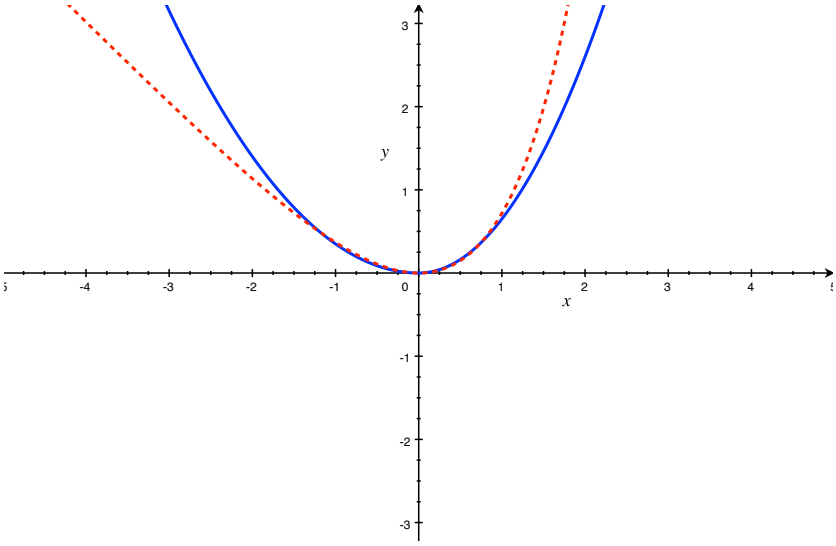

Figure A.1: Red dashed line: Linex loss with $a = 1$. Blue solid line: EKT loss with $p = 2, \alpha = 0.65$.

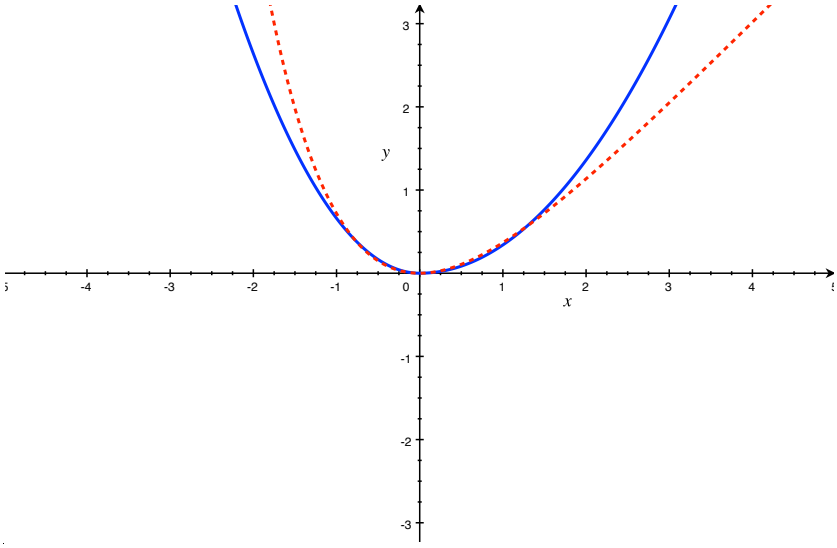

Figure A.2: Red dashed line: Linex loss with $a = -1$. Blue solid line: EKT loss with $p = 2, \alpha = 0.34$.

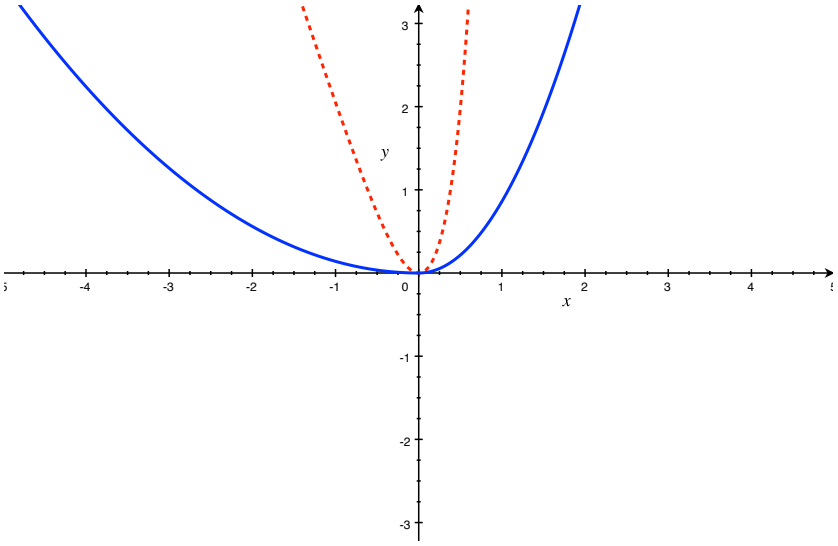

Figure A.3: Red dashed line: Linex loss with $a = 3$. Blue solid line: EKT loss with $p = 2, \alpha = 0.86$.

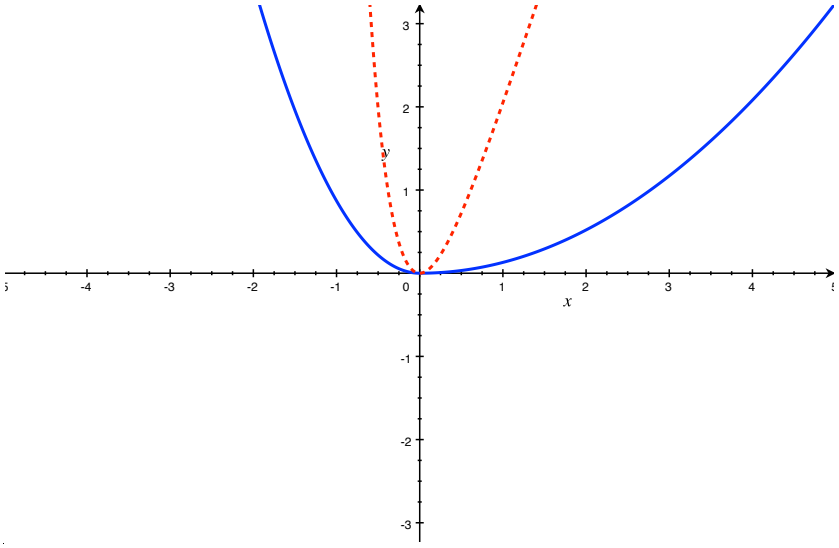

Figure A.4: Red dashed line: Linex loss with $a = -3$. Blue solid line: EKT loss with $p = 2, \alpha = 0.13$.

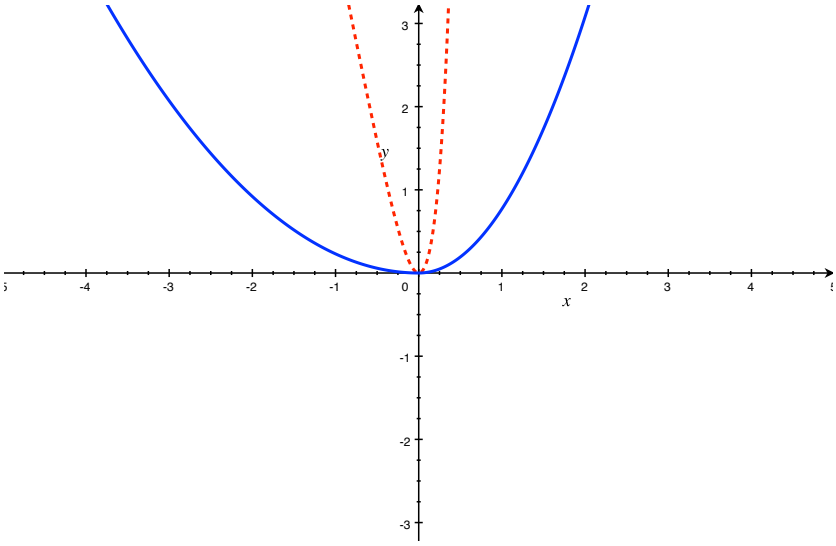

Figure A.5: Red dashed line: Linex loss with $a = 5$. Blue solid line: EKT loss with $p = 2, \alpha = 0.77$.

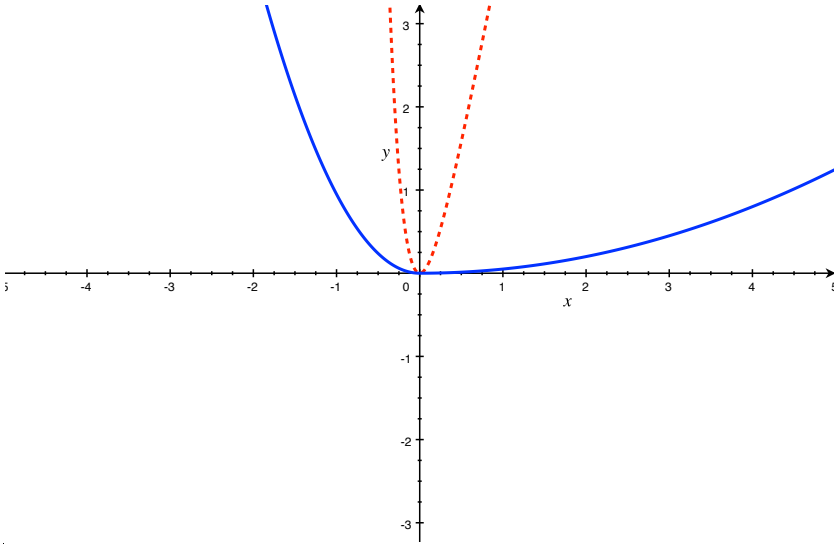

Figure A.6: Red dashed line: Linex loss with $a = -5$. Blue solid line: EKT loss with $p = 2, \alpha = 0.05$.

Appendix B

Appendix to Chapter Two

Appendix B1. Parametrization used in Simulations

Table B.1: An and Schorfheide (2007) Model Parameters and Priors - Linear Analysis

Parameters	**Names**	**Density**	**Parameter1**	**Parameter2**
τ	CRRA	Gamma	2.00	0.50
κ	Composite parameter	Gamma	0.20	0.10
ψ_1	Taylor rule infl. coeff.	Gamma	1.50	0.25
ψ_2	Taylor rule out. coeff.	Gamma	0.50	0.25
ρ_r	r_t persistence	Beta	0.50	0.20
ρ_g	g_t persistence	Beta	0.80	0.10
ρ_z	z_t persistence	Beta	0.66	0.15
r^A	St. st. interest rate	Gamma	0.50	0.50
π^A	St. st. inflation	Gamma	7.00	2.00
γ^Q	St. st. Av. output growth rate	Normal	0.40	0.20
$100\sigma_r$	ϵ_{rt} std. err.	InvGamma	0.40	4.00
$100\sigma_g$	ϵ_{gt} std. err.	InvGamma	1.00	4.00
$100\sigma_z$	ϵ_{zt} std. err.	InvGamma	0.50	4.00

NOTE: The table reports the prior distributions of the structural parameters of the An and Schorfheide model. Columns Parameter1 and Parameter2 report the means and standard deviations for the prior densities: beta, gamma, normal and the shape and the scale parameters for the inverse gamma distribution.

Table B.2: True values of the structural parameters in the DGP - Linear Analysis

Parameters	**Names**	**DGP1**
τ	CRRA	2.00
κ	Composite parameter	0.15
ψ_1	Taylor rule infl. coeff.	1.50
ψ_2	Taylor rule out. coeff.	1.00
ρ_r	r_t persistence	0.60
ρ_g	g_t persistence	0.95
ρ_z	z_t persistence	0.65
r^A	St. st. interest rate	0.40
π^A	St. st. inflation	4.00
γ^Q	St. st. Av. output growth rate	0.50
$100\sigma_r$	ϵ_{rt} std. err.	0.20
$100\sigma_g$	ϵ_{gt} std. err.	0.80
$100\sigma_z$	ϵ_{zt} std. err.	0.45

NOTE: The table lists the true values of the structural parameters in the data generating process when the DSGE model is solved by first order approximations.

Table B.3: An and Schorfheide (2007) Model Parameters and Priors - Nonlinear Analysis

Parameters	**Names**	**Density**	**Parameter1**	**Parameter2**
τ	CRRA	Gamma	2.00	0.50
κ	Composite parameter	Gamma	0.30	0.20
ψ_1	Taylor rule infl. coeff.	Gamma	1.50	0.25
ψ_2	Taylor rule out. coeff.	Gamma	0.50	0.25
ρ_r	r_t persistence	Beta	0.50	0.20
ρ_g	g_t persistence	Beta	0.80	0.10
ρ_z	z_t persistence	Beta	0.66	0.15
r^A	St. st. interest rate	Gamma	0.80	0.50
π^A	St. st. inflation	Gamma	4.00	2.00
γ^Q	St. st. Av. output growth rate	Normal	0.40	0.20
$100\sigma_r$	ϵ_{rt} std. err.	InvGamma	0.30	4.00
$100\sigma_g$	ϵ_{gt} std. err.	InvGamma	0.40	4.00
$100\sigma_z$	ϵ_{zt} std. err.	InvGamma	0.40	4.00
ν	Inv. elast. of demand	Beta	0.10	0.05
$1/g$	g - St. st. gov. spend.	Beta	0.85	0.10

NOTE: The table reports the prior distributions of the structural parameters of the An and Schorfheide model. Columns Parameter1 and Parameter2 report the means and standard deviations for the prior densities: beta, gamma, normal and the shape and the scale parameters for the inverse gamma distribution.

Table B.4: True values of the structural parameters in the DGP - Nonlinear Analysis

Parameters	**Names**	**DGP2**
τ	CRRA	2.00
κ	Composite parameter	0.33
ψ_1	Taylor rule infl. coeff.	1.50
ψ_2	Taylor rule out. coeff.	0.125
ρ_r	r_t persistence	0.75
ρ_g	g_t persistence	0.95
ρ_z	z_t persistence	0.90
r^A	St. st. interest rate	1.00
π^A	St. st. inflation	3.20
γ^Q	St. st. Av. output growth rate	0.55
$100\sigma_r$	ϵ_{rt} std. err.	0.20
$100\sigma_g$	ϵ_{gt} std. err.	0.60
$100\sigma_z$	ϵ_{zt} std. err.	0.30
ν	Inv. elast. of demand	0.10
$1/g$	g - St. st. gov. spend.	0.85

NOTE: The table lists the true values of the structural parameters in the data generating process when the DSGE model is solved by second order approximations.

Appendix B2. Grid Search Results

Table B.5: Grid Search Results - Linear Model; Objective function: log-likelihood

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
1	0.17697	75.4545	0.8333	0.1089	57.7115	0.8333	-209.7317	-208.7378	-0.9939
2	0.1155	30	0.8424	0.1596	37.1523	0.8424	-210.2507	-208.7378	-1.5129
3	0.0790	73.0303	0.8227	0.0838	77.0961	0.8227	-210.5096	-208.7378	-1.7718
4	0.1289	42.1212	0.8151	0.1385	43.9024	0.8151	-209.9893	-208.7378	-1.2514
5	0.0752	62.1212	0.8939	0.0886	72.5219	0.8939	-209.9258	-208.7378	-1.1880
6	0.1904	50	0.8575	0.1413	42.8819	0.8575	-209.7778	-208.7378	-1.0400
7	0.1558	63.9393	0.8575	0.1159	53.7948	0.8575	-209.5584	-208.7378	-0.8205
8	0.0445	41.5151	0.9303	0.0756	86.2440	0.9303	-210.6573	-208.7378	-1.9194
9	0.2	80.3030	0.9196	0.1080	58.2442	0.9196	-209.4107	-208.7378	-0.6728
10	0.04838	59.6969	0.8045	0.0688	95.4199	0.8045	-211.8043	-208.7378	-3.0664
11	0.1577	50	0.8909	0.1341	45.5783	0.8909	-209.5306	-208.7378	-0.7928
12	0.1289	45.1515	0.9363	0.1329	46.0311	0.9363	-209.4803	-208.7378	-0.7425
13	0.1846	79.6969	0.8818	0.1066	59.1420	0.8818	-209.5100	-208.7378	-0.7722
14	0.1884	51.8181	0.8515	0.1380	44.0749	0.8515	-209.7569	-208.7378	-1.0191
15	0.1481	50.6060	0.8045	0.1306	46.9755	0.8045	-209.9621	-208.7378	-1.2243
16	*0.0387*	*86.9696*	*0.8984*	*0.0502*	*133.3924*	*0.8984*	*-214.4830*	*-208.7378*	*-5.7452*
17	0.1712	65.7575	0.9045	0.1170	53.2301	0.9045	-209.3833	-208.7378	-0.6454
18	0.0906	32.4242	0.8333	0.1363	44.7023	0.8333	-209.8384	-208.7378	-1.1006
19	0.0733	87.5757	0.9136	0.0739	88.3477	0.9136	-210.7934	-208.7378	-2.0556
20	0.0906	61.5151	0.8409	0.0974	65.3420	0.8409	-209.8755	-208.7378	-1.1377

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema (minima) are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-likelihood. Columns LogPost, LogLike and LogPrior are the log posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.5: CONTINUED: Grid Search Results - Linear Model; Objective function: log-likelihood

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
21	*0.0483*	*73.6363*	*0.9227*	*0.0629*	*105.0774*	*0.9227*	*-212.0125*	*-208.7378*	*-3.2747*
22	*0.0291*	*54.8484*	*0.9393*	*0.0467*	*143.9747*	*0.9393*	*-215.4669*	*-208.7378*	*-6.7291*
23	*0.0234*	*72.4242*	*0.9409*	*0.0346*	*196.6079*	*0.9409*	*-220.7786*	*-208.7378*	*-12.0403*
24	0.1213	66.3636	0.8378	0.1050	60.1471	0.8378	-209.7487	-208.7378	-1.0109
25	0.1865	50	0.9227	0.1406	43.1523	0.9227	-209.5736	-208.7378	-0.8358
26	0.1904	85.1515	0.8333	0.1032	61.2850	0.8333	-209.8028	-208.7378	-1.0650
27	0.0886	60.3030	0.8181	0.0975	65.3062	0.8181	-210.0188	-208.7378	-1.2809
28	0.0944	71.8181	0.9136	0.0915	70.0339	0.9136	-209.7754	-208.7378	-1.0376
29	0.1213	60.30303	0.9318	0.1107	56.6910	0.9318	-209.1952	-208.7378	-0.6577
30	0.1558	50	0.8939	0.1336	45.7727	0.8939	-209.5154	-208.7378	-0.7775
31	*0.0157*	*45.7575*	*0.8287*	*0.0254*	*270.5268*	*0.8287*	*-229.1559*	*-208.7378*	*-20.4181*
32	0.1193	80.3030	0.8515	0.0940	67.9553	0.8515	-209.9104	-208.7378	-1.1725
33	0.1769	82.1212	0.85	0.1037	60.9867	0.85	-209.6980	-208.7378	-0.9602
34	0.0752	84.5454	0.8015	0.0761	85.5760	0.8015	-211.1558	-208.7378	-2.4180
35	0.1021	43.3333	0.8212	0.1238	49.9456	0.8212	-209.7911	-208.7378	-1.0533
36	0.1884	71.8181	0.8136	0.1134	55.1579	0.8136	-209.8329	-208.7378	-1.0950
37	0.0503	31.8181	0.8666	0.0939	68.1124	0.8666	-209.8400	-208.7378	-1.1022
38	0.0694	78.4848	0.8696	0.0759	85.9106	0.8696	-210.7643	-208.7378	-2.0264
39	0.1520	41.5151	0.8045	0.1474	40.7982	0.8045	-210.2240	-208.7378	-1.4862
40	0.1596	56.0606	0.8833	0.1259	48.9678	0.8833	-209.4749	-208.7378	-0.7370

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema (minima) are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-likelihood. Columns LogPost, LogLike and LogPrior are the log posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.6: Grid Search Results - Linear Model; Objective function: log-posterior

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
1	0.1251	82.1212	0.95	0.1168	52.7868	0.9217	-209.3584	-208.7447	-0.6137
2	**0.1923**	**53.6363**	**0.95**	**0.1162**	**53.1266**	**0.9483**	**-209.4196**	**-208.7447**	**-0.6749**
3	0.0272	68.1818	0.8742	0.1167	52.8424	0.9217	-209.3584	-208.7447	-0.6137
4	0.0368	81.5151	0.9121	0.1167	52.7382	0.9147	-209.3619	-208.7477	-0.6141
5	0.0810	52.4242	0.9	0.1172	52.5874	0.9045	-209.3764	-208.7448	-0.6315
6	**0.0368**	**71.8181**	**0.8393**	**0.1187**	**51.8660**	**0.8735**	**-209.4768**	**-208.7441**	**-0.7326**
7	0.1309	77.8787	0.8166	0.1168	52.8352	0.9217	-209.3584	-208.7446	-0.6137
8	0.1078	68.7878	0.9378	0.1179	52.2102	0.9358	-209.3742	-208.7459	-0.6282
9	**0.0637**	**38.4848**	**0.8045**	**0.1253**	**48.8050**	**0.8842**	**-209.4645**	**-208.7437**	**-0.7207**
10	0.1750	30	0.8439	0.1167	52.8465	0.9218	-209.3584	-208.7446	-0.6138
11	0.0810	47.5757	0.9469	0.1167	52.8373	0.9217	-209.3584	-208.7447	-0.6137
12	0.0810	82.1212	0.8106	0.1167	52.8401	0.9217	-209.3584	-208.7447	-0.6137
13	**0.1673**	**49.3939**	**0.8696**	**0.1142**	**54.1756**	**0.8871**	**-209.4263**	**-208.7451**	**-0.6812**
14	0.0253	76.0606	0.8530	0.1167	52.8367	0.9217	-209.3584	-208.7447	-0.6137
15	0.0906	83.3333	0.8666	0.1168	52.8285	0.9218	-209.3584	-208.7446	-0.6137
16	**0.1021**	**88.1818**	**0.8181**	**0.1216**	**50.4381**	**0.8731**	**-209.4876**	**-208.7444**	**-0.7431**
17	*0.1520*	*83.9393*	*0.9378*	*0.3790*	*11.9372*	*0.7921*	*-220.5359*	*-208.8073*	*-11.7285*
18	0.1002	74.2424	0.8984	0.1167	52.8398	0.9218	-209.3584	-208.7447	-0.6137
19	0.0407	71.2121	0.9106	0.1171	52.5367	0.9125	-209.3642	-208.7476	-0.6166
20	0.0483	71.2121	0.9439	0.1168	52.7854	0.9422	-209.3925	-208.7447	-0.6477

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.6: CONTINUED: Grid Search Results - Linear Model; Objective function: log-posterior

ID	Init ν	Init ϕ	Init 1/g	ν	ϕ	1/g	LogPost	LogLike	LogPrior
21	**0.1347**	**77.8787**	**0.8803**	**0.0991**	**63.5734**	**0.8804**	**-209.6225**	**-208.7426**	**-0.8798**
22	0.0387	53.6363	0.95	0.1169	52.7839	0.9482	-209.4189	-208.7444	-0.6745
23	0.1846	54.8484	0.8651	0.1168	52.8230	0.9217	-209.3584	-208.7447	-0.6137
24	0.1635	51.2121	0.9318	0.1169	52.7698	0.9217	-209.3584	-208.7445	-0.6139
25	**0.1193**	**47.5757**	**0.8772**	**0.1127**	**54.9832**	**0.8952**	**-209.4061**	**-208.7449**	**-0.6612**
26	0.0330	71.2121	0.9469	0.1168	52.8233	0.9218	-209.3584	-208.7447	-0.6137
27	0.0176	40.9090	0.8712	0.1167	52.8422	0.9218	-209.3584	-208.7446	-0.6137
28	**0.0598**	**77.2727**	**0.8303**	**0.1193**	**51.5660**	**0.8711**	**-209.4884**	**-208.7444**	**-0.7439**
29	**0.1501**	**31.8181**	**0.8772**	**0.1159**	**53.2749**	**0.8895**	**-209.4156**	**-208.7447**	**-0.6709**
30	0.1923	60.3030	0.9045	0.1168	52.8355	0.9217	-209.3584	-208.7447	-0.6137
31	0.1616	37.8787	0.9151	0.1167	52.8452	0.9161	-209.3605	-208.7447	-0.6158
32	**0.0771**	**75.4545**	**0.9151**	**0.1164**	**53.0296**	**0.9163**	**-209.3604**	**-208.7449**	**-0.6155**
33	0.0426	30	0.8	0.1167	52.8716	0.9217	-209.3584	-208.7447	-0.6137
34	**0.0157**	**59.696**	**0.9363**	**0.1163**	**53.0658**	**0.9999**	**-215.3805**	**-208.7446**	**-6.6359**
35	0.0311	37.8787	0.9181	0.1168	52.8071	0.9187	-209.3590	-208.7447	-0.6143
36	0.1654	86.3636	0.8121	0.1168	52.8339	0.9217	-209.3584	-208.7447	-0.6137
37	0.0272	33.0303	0.9106	0.1168	52.8317	0.9217	-209.3584	-208.7447	-0.6137
38	**0.2**	**42.1212**	**0.8727**	**0.1155**	**53.4998**	**0.8871**	**-209.4239**	**-208.7446**	**-0.6792**
39	*0.0407*	*63.3333*	*0.8984*	*0.0599*	*107.8570*	*0.8989*	*-212.3192*	*-208.7797*	*-3.5395*
40	0.0291	87.5757	0.8015	0.1168	52.8356	0.9217	-209.3584	-208.7447	-0.6137

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.7: Grid Search Results - Linear Model, Correct Specification - Output Growth

ID	τ	ν	k	$1/g$	ψ_1	ψ_2	ρ_r	ρ_g	ρ_z
1	**1.0442**	**0.1**	**0.1209**	**0.85**	**1.5408**	**0.8401**	**0.4161**	**0.9578**	**0.7634**
2	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
3	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
4	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
5	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
6	**1.0442**	**0.1**	**0.1209**	**0.85**	**1.5408**	**0.8401**	**0.4161**	**0.9578**	**0.7634**
7	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
8	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
9	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
10	**1.0442**	**0.1**	**0.1209**	**0.85**	**1.5408**	**0.8401**	**0.4161**	**0.9578**	**0.7634**
11	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
12	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
13	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
14	**1.0442**	**0.1**	**0.1209**	**0.85**	**1.5408**	**0.8401**	**0.4161**	**0.9578**	**0.7634**
15	**1.0442**	**0.1**	**0.1209**	**0.85**	**1.5408**	**0.8401**	**0.4161**	**0.9578**	**0.7634**
16	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
17	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
18	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
19	1.5632	0.1	0.1376	0.85	1.4517	0.8996	0.5158	0.9421	0.6008
20	**1.0442**	**0.1**	**0.1209**	**0.85**	**1.5408**	**0.8401**	**0.4161**	**0.9578**	**0.7634**

NOTE: The table reproduces 20 iterations form the grid search results for the setting $M_2(L)/D_2(L)$ from An and Schorfheide (2007). The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.7: CONTINUED: Grid Search Results - Linear Model, Correct Specification - Output Growth

ID	r^A	π^A	γ^Q	$100 \times \sigma_r$	$100 \times \sigma_g$	$100 \times \sigma_z$	**LogPost**	**LogLike**	**LogPrior**
1	**0.2292**	**3.4983**	**0.4093**	**0.2322**	**0.0701**	**0.3035**	**-144.8601**	**-156.2941**	**11.4340**
2	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
3	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6175
4	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6175
5	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
6	**0.2292**	**3.4983**	**0.4093**	**0.2322**	**0.0701**	**0.3035**	**-144.8601**	**-156.2940**	**11.4339**
7	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1241	13.6174
8	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
9	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
10	**0.2292**	**3.4983**	**0.4093**	**0.2322**	**0.0701**	**0.3035**	**-144.8601**	**-156.2942**	**11.4341**
11	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
12	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1241	13.6174
13	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
14	**0.2292**	**3.4983**	**0.4093**	**0.2322**	**0.0701**	**0.3035**	**-144.8601**	**-156.2941**	**11.4340**
15	**0.2292**	**3.4983**	**0.4093**	**0.2322**	**0.0701**	**0.3035**	**-144.8601**	**-156.2941**	**11.4340**
16	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
17	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6174
18	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6175
19	0.1691	3.4439	0.3893	0.2116	0.7044	0.4866	-138.5067	-152.1242	13.6175
20	**0.2292**	**3.4983**	**0.4093**	**0.2322**	**0.0701**	**0.3035**	**-144.8601**	**-156.2940**	**11.4339**

NOTE: The table reproduces 20 iterations form the grid search results for the setting $M_2(L)/D_2(L)$ from An and Schorfheide (2007). The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.8: Grid Search Results - Linear Model, Misspecification - Output Growth

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
1	0.0925	40.3030	0.8030	0.2467	81.1004	0.8030	-394.3478	-388.9316	-5.4162
2	0.081	83.9393	0.9287	0.1559	143.7475	0.9287	-396.3188	-388.9316	-7.3872
3	0.1462	40.3030	0.85	0.3047	60.5880	0.85	-394.9298	-388.9316	-5.9982
4	0.1654	73.6363	0.8257	0.2318	88.0033	0.8257	-394.1838	-388.9316	-5.2521
5	0.1078	61.5151	0.8151	0.2145	97.2482	0.8151	-394.4003	-388.9316	-5.4687
6	0.0618	86.9696	0.9363	0.1280	180.8913	0.9363	-399.2683	-388.9316	-10.3367
7	0.1654	79.0909	0.8121	0.2229	92.6076	0.8121	-394.3286	-388.9316	-5.3970
8	0.1328	36.0606	0.9484	0.3102	59.0572	0.9484	-394.8982	-388.9316	-5.9667
9	0.1522	58.4848	0.8878	0.2549	77.6375	0.8878	-393.9230	-388.9316	-4.9914
10	0.1846	71.8181	0.9090	0.2435	82.5034	0.9090	-393.8119	-388.9316	-4.8802
11	0.1673	81.5151	0.9181	0.2200	94.1561	0.9181	-393.8861	-388.9316	-4.9545
12	0.1846	68.1818	0.8378	0.2505	79.4404	0.8378	-394.1335	-388.9316	-5.2019
13	0.1385	42.1212	0.8469	0.2919	64.4355	0.8469	-394.6522	-388.9316	-5.7206
14	0.0829	77.2727	0.8545	0.1651	134.3083	0.8545	-395.8896	-388.9316	-6.9580
15	0.2	40.9090	0.9181	0.3356	52.5846	0.9181	-395.5842	-388.9316	-6.6525
16	0.0925	74.8484	0.8045	0.1786	122.0974	0.8045	-395.4825	-388.9316	-6.5509
17	0.1002	43.9393	0.8590	0.2462	81.2875	0.8590	-393.9950	-388.9316	-5.0633
18	0.1827	62.7272	0.8651	0.2613	75.0776	0.8651	-394.0731	-388.9316	-5.1415
19	0.0464	49.3939	0.9333	0.1202	194.3027	0.9333	-400.4488	-388.9316	-11.5171
20	0.1808	55.4545	0.85	0.2782	68.9114	0.85	-394.3785	-388.9316	-5.4468

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-likelihood. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.8: CONTINUED: Grid Search Results - Linear Model, Misspecification - Output Growth

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
21	0.0982	30	0.9378	0.2990	62.2634	0.9378	-394.5804	-388.9316	-5.6488
22	0.0829	47.5757	0.9106	0.2124	98.4787	0.9106	-393.9896	-388.9316	-5.0580
23	0.1731	59.6969	0.8060	0.2637	74.1373	0.8060	-394.4582	-388.9316	-5.5266
24	0.0982	51.8181	0.9015	0.2238	92.1131	0.9015	-393.8750	-388.9316	-4.9433
25	0.1712	76.6666	0.9469	0.2293	89.2545	0.9469	-393.8708	-388.9316	-4.9383
26	0.0886	38.4848	0.9196	0.2470	80.9678	0.9196	-393.8140	-388.9316	-4.8824
27	0.1289	79.6969	0.8045	0.2023	104.7096	0.8045	-394.6964	-388.9316	-5.7648
28	0.1385	33.0303	0.9106	0.3294	54.0657	0.9106	-395.3950	-388.9316	-6.4634
29	0.0234	39.0909	0.9242	0.0038	0.01218	0.9242	-582.2639	-539.4220	-42.8418
30	0.0771	76.6666	0.9333	0.1586	140.8666	0.9333	-396.1259	-388.9316	-7.1943
31	0.1865	51.2121	0.8378	0.2929	64.0984	0.8378	-394.7281	-388.9316	-5.7965
32	0.1347	53.0303	0.8969	0.2562	77.0983	0.8969	-393.9065	-388.9316	-4.9749
33	0.1174	83.9393	0.8318	0.1890	113.9584	0.8318	-394.8883	-388.9316	-5.9566
34	0.1289	71.8181	0.8121	0.2141	97.5005	0.8121	-394.4274	-388.9316	-5.4958
35	0.1462	46.9696	0.8621	0.2819	67.6664	0.8621	-394.3785	-388.9316	-5.4468
36	0.1788	60.9090	0.9015	0.2637	74.1553	0.9015	-393.9685	-388.9316	-5.0369
37	0.1616	54.8484	0.8803	0.2695	71.9700	0.8803	-394.1061	-388.9316	-5.1745
38	0.1923	71.2121	0.9303	0.2477	80.6566	0.9303	-393.8221	-388.9316	-4.8905
39	0.1347	39.6969	0.8803	0.2974	62.7386	0.8803	-394.6134	-388.9316	-5.6818
40	0.1769	63.3333	0.9136	0.2573	76.6527	0.9136	-393.8849	-388.9316	-4.9533

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-likelihood. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.9: Grid Search Results - Linear Model, Misspecification - Inflation

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
1	0.1251	35.4545	0.85	0.1648	42.9531	0.85	-209.8275	-208.2844	-1.5430
2	0.1174	42.7272	0.9287	0.1464	49.4403	0.9287	-209.3011	-208.2844	-1.0167
3	0.0157	73.6363	0.9318	0.0238	346.5420	0.9318	-236.6616	-208.2844	-28.3771
4	0.1059	30	0.9378	0.1699	41.5565	0.9378	-209.7246	-208.2844	-1.4392
5	0.1443	39.6969	0.8	0.1650	42.9117	0.8	-210.1557	-208.2844	-1.8712
6	0.1117	74.2424	0.8272	0.1060	71.5111	0.8272	-209.8837	-208.2844	-1.5993
7	0.1808	31.2121	0.8196	0.2036	33.1646	0.8196	-211.0690	-208.2844	-2.7846
8	0.0464	73.0303	0.8257	0.0663	119.2492	0.8257	-213.0235	-208.2844	-4.7390
9	0.0195	50	0.8954	0.0335	243.9272	0.8954	-225.1621	-208.2844	-16.8777
10	0.0579	80.9090	0.8803	0.0736	106.7029	0.8803	-211.7289	-208.2844	-3.4445
11	0.0963	64.5454	0.8469	0.1075	70.3397	0.8469	-209.7218	-208.2844	-1.4373
12	0.0483	67.5757	0.9090	0.0706	111.4706	0.9090	-212.0211	-208.2844	-3.7367
13	0.0675	88.1818	0.9409	0.0772	101.3202	0.9409	-211.2671	-208.2844	-2.9827
14	0.1213	88.1818	0.8878	0.0993	76.8624	0.8878	-209.8038	-208.2844	-1.5193
15	0.0656	68.7878	0.9333	0.0856	90.5797	0.9333	-210.5142	-208.2844	-2.2297
16	0.1596	86.3636	0.9106	0.1091	69.2028	0.9106	-209.4415	-208.2844	-1.1571
17	0.1539	46.9696	0.8939	0.1532	46.8522	0.8939	-209.4313	-208.2844	-1.1468
18	0.0771	37.8787	0.95	0.1267	58.4343	0.95	-209.2803	-208.2844	-0.9958
19	0.1097	79.6969	0.9454	0.1013	75.1890	0.9454	-209.7128	-208.2844	-1.4283
20	0.1462	57.2727	0.9378	0.1346	54.5209	0.9378	-209.2305	-208.2844	-0.9464

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-likelihood. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.9: CONTINUED: Grid Search Results - Linear Model, Misspecification - Inflation

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
21	0.0675	50	0.8272	0.1017	74.8867	0.8272	-210.0208	-208.2844	-1.7363
22	0.1769	82.1212	0.8530	0.1154	64.9711	0.8530	-209.5278	-208.2844	-1.2434
23	0.1769	75.4545	0.8333	0.1213	61.4241	0.8333	-209.5722	-208.2844	-1.2877
24	0.1155	30	0.8424	0.1764	39.5677	0.8424	-210.1351	-208.2844	-1.8507
25	0.0790	73.0303	0.8227	0.0919	83.7332	0.8227	-210.4926	-208.2844	-2.2082
26	0.1289	42.1212	0.8151	0.1532	46.8640	0.8151	-209.8364	-208.2844	-1.5520
27	0.0752	62.1212	0.8939	0.0972	78.7631	0.8939	-209.8762	-208.2844	-1.5918
28	0.1904	50	0.8575	0.1574	45.3843	0.8575	-209.6469	-208.2844	-1.3625
29	0.1558	63.9393	0.8575	0.1288	57.3113	0.8575	-209.3968	-208.2844	-1.1124
30	0.0445	41.5151	0.9303	0.0819	95.0046	0.9303	-210.7996	-208.2844	-2.5152
31	0.2	80.3030	0.9196	0.1205	61.8925	0.9196	-209.2500	-208.2844	-0.9656
32	0.0483	59.6967	0.8045	0.0746	105.172	0.8045	-212.0475	-208.2844	-3.7631
33	0.1577	50	0.8909	0.1490	48.4225	0.8909	-209.3815	-208.2844	-1.0971
34	0.1289	45.1515	0.9363	0.1473	49.0798	0.9363	-209.3244	-208.2844	-1.0400
35	0.1846	79.6969	0.8818	0.1187	62.9350	0.8818	-209.3520	-208.2844	-1.0675
36	0.1884	51.8181	0.8515	0.1537	46.6753	0.8515	-209.6191	-208.2844	-1.3347
37	0.1481	50.6060	0.8045	0.1450	49.9819	0.8045	-209.8073	-208.2844	-1.5228
38	0.0387	86.9696	0.8984	0.0539	148.7213	0.8984	-215.3164	-208.2844	-7.0319
39	0.1712	65.7575	0.9045	0.1302	56.6103	0.9045	-209.2216	-208.2844	-0.9371
40	0.0906	32.4242	0.8333	0.1503	47.9080	0.8333	-209.6766	-208.2844	-1.3922

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by first-order accurate solutions. To save space, only results for ν, ϕ and $1/g$ are reported here. Recall that the other parameters were fixed to their true values. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extrema are given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-likelihood. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode. The sample size is N=80.

Table B.10: Grid Search Results - Nonlinear Model, N=80

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
1	0.0272	65.7575	0.8439	0.0893	45.5915	0.9217	-363.2130	-380.4883	17.2753
2	0.2	60.9090	0.8803	0.0895	45.5215	0.9215	-363.2069	-380.4790	17.2720
3	0.1251	76.6766	0.8090	0.0926	44.1177	0.9204	-363.2710	-380.5384	17.2667
4	0.1884	45.1515	0.9378	0.0893	45.5943	0.9217	-363.2187	-380.4946	17.2759
5	0.1309	88.1818	0.9090	0.0893	45.5892	0.9218	-363.2149	-380.4891	17.2741
6	0.1769	70.6060	0.9242	0.0893	45.6063	0.9218	-363.2108	-380.4855	17.2746
7	0.0311	55.4545	0.9075	0.0893	45.5941	0.9218	-363.2755	-380.5509	17.2753
8	0.1232	41.5151	0.9196	0.0893	45.8066	0.9209	-363.3062	-380.5900	17.2837
9	0.0522	74.8484	0.8378	0.0855	46.0131	0.8836	-363.1648	-380.2760	17.1111
10	0.1923	88.7878	0.8227	0.0887	45.7950	0.9109	-363.1430	-380.4035	17.2604
11	0.1443	63.3333	0.8787	0.0914	47.8495	0.8891	-364.4250	-380.8295	16.4044
12	0.1251	68.7878	0.8090	0.0893	45.5744	0.9217	-363.2038	-380.4789	17.2750
13	0.1059	62.1212	0.9106	0.0893	45.5911	0.9217	-363.2132	-380.4884	17.2751
14	0.1827	75.4545	0.8545	0.0895	45.5046	0.9157	-363.2681	-380.5438	17.2756
15	0.1654	49.3939	0.8560	0.0894	45.6021	0.9293	-363.2288	-380.4977	17.2689
16	0.2	75.4545	0.8181	0.0888	45.3135	0.9124	-363.0732	-380.3313	17.2581
17	0.0483	34.8484	0.8045	0.0893	45.5936	0.9217	-363.213	-380.4854	17.2724
18	0.1002	82.7272	0.8106	0.0893	45.5897	0.9217	-363.2132	-380.4883	17.2751
19	0.0253	58.4848	0.8772	0.0888	45.7308	0.9176	-363.1433	-380.3959	17.2526

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by second-order accurate solutions. The sample size is N=80. None of the parameters are fixed in the nonlinear approximations but to save space only results for ν, ϕ and $1/g$ are reported here. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extremum for each iteration is given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode.

Table B.10: CONTINUED: Grid Search Results - Nonlinear Model, N=80

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
20	0.0925	78.4848	0.9166	0.0892	45.7590	0.9200	-363.3347	-380.6111	17.2763
21	0.0483	84.5454	0.8242	0.0894	45.6094	0.9200	-363.2348	-380.5051	17.2703
22	0.1558	68.7878	0.9469	0.0894	45.5557	0.9217	-363.2133	-380.4885	17.2751
23	0.0790	41.5151	0.8636	0.0894	45.5584	0.9217	-363.2146	-380.4898	17.2751
24	0.0157	39.0909	0.9454	0.0893	45.5911	0.9217	-363.2134	-380.4885	17.2751
25	0.0426	31.2121	0.8939	0.0893	45.5858	0.9218	-363.2166	-380.4917	17.2750
26	0.0618	43.9393	0.8681	0.0895	45.5554	0.9212	-363.2424	-380.5193	17.2768
27	0.0656	57.2727	0.9454	0.0960	43.2809	0.9383	-363.3837	-380.5569	17.1732
28	0.1040	40.9090	0.8969	0.0895	45.4345	0.9205	-363.2064	-380.4845	17.2781
29	0.1596	90	0.8757	0.0907	44.9594	0.9158	-363.2452	-380.5096	17.2644
30	0.1980	76.6666	0.9454	0.0893	45.5843	0.9306	-363.2162	-380.4860	17.2695
31	0.0138	30.6060	0.8666	0.0895	45.5136	0.9144	-363.2200	-380.4930	17.2729
32	0.0656	83.3333	0.8469	0.0895	45.6528	0.9203	-363.2934	-380.5724	17.2790
33	0.1673	90	0.8454	0.0892	45.6662	0.9217	-363.2092	-380.4838	17.2746
34	0.0848	73.6363	0.8742	0.0893	45.5910	0.9217	-363.2143	-380.4895	17.2752
35	0.1827	40.9090	0.9303	0.0882	46.1767	0.9281	-363.1976	-380.4643	17.2667
36	0.0771	70	0.8772	0.0901	44.8376	0.8996	-363.3267	-380.5690	17.2422
37	0.0714	86.3636	0.9378	0.0884	46.1985	0.9240	-363.2056	-380.4749	17.2693
38	0.1481	85.7575	0.8939	0.0892	45.6290	0.9120	-363.1888	-380.4551	17.2663
39	0.0311	64.5454	0.8484	0.0893	45.6193	0.9216	-363.2205	-380.4968	17.2762
40	0.0944	31.2121	0.8166	0.0889	45.7825	0.9234	-363.2157	-380.4919	17.2762

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by second-order accurate solutions. The sample size is N=80. None of the parameters are fixed in the nonlinear approximations but to save space only results for ν, ϕ and $1/g$ are reported here. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extremum for each iteration is given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode.

Table B.11: Grid Search Results - Nonlinear Model, N=3000

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
1	0.1078	63.9393	0.8	0.1150	52.3800	0.9218	-15037.732	-15053.332	15.600
2	0.1923	69.3939	0.8045	0.1151	52.3493	0.9217	-15037.734	-15053.334	15.600
3	0.1712	39.6969	0.8166	0.1150	52.3747	0.9217	-15037.737	-15053.338	15.600
4	0.0215	87.5757	0.8742	0.1150	52.3755	0.9217	-15037.736	-15053.337	15.600
5	0.0157	74.8484	0.8227	0.1148	52.4689	0.9142	-15037.435	-15053.026	15.591
6	0.0483	73.0303	0.8666	0.1150	52.3812	0.9217	-15037.737	-15053.337	15.599
7	0.1174	40.3030	0.8772	0.1150	52.3788	0.9218	-15037.733	-15053.334	15.600
8	0.0311	55.4545	0.8924	0.1150	52.3750	0.9217	-15037.742	-15053.343	15.600
9	0.1155	68.1818	0.9484	0.1150	52.3787	0.9217	-15037.749	-15053.350	15.600
10	0.0982	56.0606	0.9166	0.1149	52.4312	0.9185	-15037.732	-15053.331	15.599
11	0.1462	48.1818	0.8393	0.1150	52.3800	0.9217	-15037.734	-15053.334	15.600
12	0.0867	75.4545	0.8727	0.1151	52.3578	0.9152	-15037.727	-15053.325	15.597
13	0.1366	30.6060	0.8439	0.1150	52.3800	0.9217	-15037.729	-15053.329	15.600
14	0.0925	80.3030	0.8484	0.1150	52.3836	0.9218	-15037.742	-15053.342	15.600
15	0.1002	45.1515	0.8454	0.1150	52.3827	0.9217	-15037.729	-15053.330	15.600
16	0.0886	59.6969	0.9424	0.1148	52.4563	0.9205	-15037.204	-15052.801	15.596
17	0.0483	89.3939	0.8924	0.1150	52.3789	0.9217	-15037.729	-15053.330	15.600
18	0.0637	79.6969	0.8515	0.1150	52.3727	0.9217	-15037.734	-15053.335	15.600
19	0.0848	64.5454	0.8363	0.1150	52.3757	0.9217	-15037.721	-15053.321	15.600
20	0.1328	82.1212	0.8909	0.1150	52.4015	0.9125	-15037.737	-15053.332	15.595

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by second-order accurate solutions. The sample size is N=3000. None of the parameters are fixed in the nonlinear approximations but to save space only results for ν, ϕ and $1/g$ are reported here. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extremum for each iteration is given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode.

Table B.11: CONTINUED: Grid Search Results - Nonlinear Model, N=3000

ID	**Init ν**	**Init ϕ**	**Init 1/g**	ν	ϕ	**1/g**	**LogPost**	**LogLike**	**LogPrior**
21	0.2	51.8181	0.8833	0.1150	52.3804	0.9217	-15037.736	-15053.337	15.600
22	0.0291	37.8787	0.9287	0.1150	52.4068	0.9266	-15037.731	-15053.329	15.598
23	0.01	79.6967	0.9363	0.1150	52.3827	0.9311	-15037.677	-15053.270	15.593
24	0.1865	57.2727	0.8954	0.1161	51.8308	0.9137	-15037.740	-15053.334	15.594
25	0.1788	47.5757	0.8439	0.1151	52.3524	0.9209	-15037.712	-15053.311	15.599
26	0.1731	30	0.95	0.1150	52.3877	0.9217	-15037.843	-15053.443	15.600
27	0.1117	84.5454	0.8712	0.1150	52.3938	0.9253	-15037.705	-15053.305	15.599
28	0.1347	42.7272	0.8121	0.1150	52.3755	0.9217	-15037.703	-15053.304	15.600
29	0.1923	50	0.8227	0.1169	51.4227	0.9192	-15037.973	-15053.571	15.598
30	0.01	56.0606	0.8015	0.1150	52.3759	0.9217	-15037.733	-15053.334	15.600
31	0.1462	79.6969	0.9030	0.1145	52.6322	0.9131	-15037.720	-15053.314	15.594
32	0.1616	37.8787	0.8045	0.1150	52.3809	0.9217	-15037.732	-15053.332	15.600
33	0.1385	68.7878	0.9060	0.1144	52.6854	0.914	-15037.749	-15053.345	15.596
34	0.0522	84.5454	0.8560	0.1150	52.3620	0.9177	-15037.746	-15053.345	15.599
35	0.1174	46.3636	0.8227	0.1150	52.3774	0.9217	-15037.735	-15053.335	15.600
36	0.1788	50	0.9196	0.1150	52.3775	0.9210	-15037.730	-15053.330	15.603
37	0.1904	77.8787	0.8196	0.1150	52.3723	0.9197	-15037.740	-15053.340	15.600
38	0.0445	46.9696	0.8075	0.1150	52.3808	0.9218	-15037.731	-15053.331	15.600
39	0.0464	60.9090	0.9121	0.1151	52.3498	0.9172	-15037.769	-15053.368	15.599
40	0.0311	37.8787	0.9424	0.1150	52.3824	0.9217	-15037.730	-15053.330	15.600

NOTE: The table reproduces 40 iterations from the grid search results when the model is solved by second-order accurate solutions. The sample size is N=3000. None of the parameters are fixed in the nonlinear approximations but to save space only results for ν, ϕ and $1/g$ are reported here. The column ID is the identifier of the search. Init ν, Init ϕ and Init $1/g$ list the starting grid points used for the optimization of the objective function. The obtained global extremum for each iteration is given in columns ν, ϕ and $1/g$. The objective function that is optimized is the log-posterior. Columns LogPost, LogLike and LogPrior are the log-posterior, log-likelihood and the log-prior evaluated at the obtained mode.

Appendix B3. System Matrices

For the parameter ID 6 in Table 2.2, the (7x7) matrix A is:

$$A_{1,1} = \begin{pmatrix} 0 & 0 & -1 & -0.341786 & 0 & 0.95 & 0.255142 \\ 0 & 0 & 0 & -0.151199 & 0 & 0 & 0.125646 \\ 0 & 0 & 0 & -0.341786 & 0 & 0.95 & 0.255142 \\ 0 & 0 & 0 & 0.372565 & 0 & 0 & 0.177444 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{1,1} = \begin{pmatrix} -0.001139 & 0.008 & 0.001766 \\ -0.000503 & 0 & 0.000869 \\ -0.001139 & 0.008 & 0.001766 \\ 0.001241 & 0 & 0.001228 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the parameter ID 7 in Table 2.2, the (7x7) matrix A is:

$$A_{1,2} = \begin{pmatrix} 0 & 0 & -1 & -0.341786 & 0 & 0.95 & 0.255141 \\ 0 & 0 & 0 & -0.151201 & 0 & 0 & 0.125647 \\ 0 & 0 & 0 & -0.341786 & 0 & 0.95 & 0.255141 \\ 0 & 0 & 0 & 0.372564 & 0 & 0 & 0.177445 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{1,2} = \begin{pmatrix} -0.001139 & 0.008 & 0.001766 \\ -0.000503 & 0 & 0.000869 \\ -0.001139 & 0.008 & 0.001766 \\ 0.001241 & 0 & 0.001228 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the random point from column three of Table 2.2, the (7x7) matrix

A is:

$$A_{1,3} = \begin{pmatrix} 0 & 0 & -1 & -0.354682 & 0 & 0.95 & 0.265288 \\ 0 & 0 & 0 & -0.115770 & 0 & 0 & 0.096888 \\ 0 & 0 & 0 & -0.354682 & 0 & 0.95 & 0.265288 \\ 0 & 0 & 0 & 0.388664 & 0 & 0 & 0.164248 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{1,3} = \begin{pmatrix} -0.001182 & 0.008 & 0.001836 \\ -0.000385 & 0 & 0.000670 \\ -0.001182 & 0.008 & 0.001836 \\ 0.001295 & 0 & 0.001137 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the parameter ID 12 in Table 2.4, the (7x7) matrix A is:

$$A_{2,1} = \begin{pmatrix} 0 & 0 & -0.612077 & -0.581883 & 0 & 0.634741 & -0.123886 \\ 0 & 0 & 0.088169 & -0.132253 & 0 & -0.007017 & -0.111330 \\ 0 & 0 & 0.387922 & -0.581883 & 0 & 0.634741 & -0.123886 \\ 0 & 0 & -0.191929 & 0.287894 & 0 & 0.249686 & 0.143646 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{2,1} = \begin{pmatrix} -0.001939 & 0.005345 & -0.000857 \\ -0.000440 & 0 & -0.000770 \\ -0.001939 & 0.005345 & -0.000857 \\ 0.000959 & 0.002102 & 0.000994 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the parameter ID 15 in Table 2.4, the (7x7) matrix A is:

$$A_{2,2} = \begin{pmatrix} 0 & 0 & -0.612055 & -0.581916 & 0 & 0.634730 & -0.123915 \\ 0 & 0 & 0.088146 & -0.132220 & 0 & -0.006990 & -0.111312 \\ 0 & 0 & 0.387944 & -0.581916 & 0 & 0.634730 & -0.123915 \\ 0 & 0 & -0.191934 & 0.287901 & 0 & 0.249692 & 0.143646 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{2,2} = \begin{pmatrix} -0.001939 & 0.005345 & -0.000857 \\ -0.000440 & 0 & -0.000770 \\ -0.001939 & 0.005345 & -0.000857 \\ 0.000959 & 0.002102 & 0.000994 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the parameter random point from Table 2.4, the (7x7) matrix A is:

$$A_{2,3} = \begin{pmatrix} 0 & 0 & 0.231345 & -0.347018 & 0 & -0.12700 & -0.184072 \\ 0 & 0 & 0.268812 & -0.403219 & 0 & 0.719533 & -0.024243 \\ 0 & 0 & -0.153667 & 0.230501 & 0 & 0.2116070.139859 & \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{2,3} = \begin{pmatrix} -0.001344 & 0.006059 & -0.0001 \\ -0.001156 & -0.001069 & -0.001274 \\ -0.001344 & 0.006059 & -0.00016 \\ 0.000768 & 0.001781 & 0.00096 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the parameter ID 6 from Table 2.5, the (7x7) matrix A is:

$$A_{3,1} = \begin{pmatrix} 0 & 0 & -1 & -0.393211 & 0 & 0.95 & 0.295650 \\ 0 & 0 & 0 & -0.152827 & 0- & 0 & 0.127884 \\ 0 & 0 & 0 & -0.393211 & 0 & 0.95 & 0.295650 \\ 0 & 0 & 0 & 0.405529 & 0 & 0 & 0.154012 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{3,1} = \begin{pmatrix} -0.001310 & 0.008 & 0.002046 \\ -0.000509 & 0 & 0.000885 \\ -0.001310 & 0.008 & 0.002046 \\ 0.001351 & 0 & 0.001066 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the parameter ID 7 from Table 2.5, the (7x7) matrix A is:

$$A_{3,2} = \begin{pmatrix} 0 & 0 & -1 & -0.393211 & 0 & 0.95 & 0.295650 \\ 0 & 0 & 0 & -0.152834 & 0- & 0 & 0.127890 \\ 0 & 0 & 0 & -0.393211 & 0 & 0.95 & 0.295650 \\ 0 & 0 & 0 & 0.405528 & 0 & 0 & 0.154014 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{3,2} = \begin{pmatrix} -0.001310 & 0.008 & 0.002046 \\ -0.000509 & 0 & 0.000885 \\ -0.001310 & 0.008 & 0.002046 \\ 0.001351 & 0 & 0.001066 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

For the random point from Table 2.5, the (7x7) matrix A is:

$$A_{3,3} = \begin{pmatrix} 0 & 0 & -1 & -0.392873 & 0 & 0.95 & 0.295302 \\ 0 & 0 & 0 & -0.177349 & 0- & 0 & 0.147872 \\ 0 & 0 & 0 & -0.392873 & 0 & 0.95 & 0.295302 \\ 0 & 0 & 0 & 0.400259 & 0 & 0 & 0.158884 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.95 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.65 \end{pmatrix}$$

and the (7x3) matrix B is:

$$B_{3,3} = \begin{pmatrix} -0.001309 & 0.008 & 0.002044 \\ -0.000591 & 0 & 0.001023 \\ -0.001309 & 0.008 & 0.002044 \\ 0.001334 & 0 & 0.001099 \\ 0.002 & 0 & 0 \\ 0 & 0.008 & 0 \\ 0 & 0 & 0.0045 \end{pmatrix}$$

Appendix B4. Estimation

Frequentist Approach

In a frequentist framework, a DSGE model is commonly estimated by maximum likelihood. The likelihood of the model $p(Y|\theta)$ is evaluated with the Kalman filter when the DSGE is linearized or with the particle filter when the DSGE model is nonlinear. The optimization of the objective function is frequently performed using the minimisation routine CSMINWELL by Sims. The algorithm finds the global minimum, being robust to 'cliffs' of the likelihood.

Bayesian Approach

In a Bayesian framework, the posterior moments of the DSGE model's structural parameters are commonly computed with numerical methods such as the Random Walk Metropolis (RWM). Below is a description of how the RWM algorithm generates draws from the posterior distribution of θ.

First, the prior density $p(\theta)$ is combined with $p(Y|\theta)$, the conditional density of the data Y given the parameters. The likelihood $p(Y|\theta)$ can be evaluated with a Kalman filter when the DSGE is linearized or with the particle filter when the DSGE model is nonlinear. Then, according to Bayes theorem, the posterior distribution, $p(\theta|Y)$ is given by:

$$p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)},$$

where $p(Y) = \int p(Y|\theta)p(\theta)d\theta$ is the marginal likelihood or data density.

The Random-Walk Metropolis algorithm for DSGE models (see e.g. Schorfheide [2000]) takes the following form:

1. Maximize the log-posterior, which up to a constant is given by $\ln p(Y|\theta) + \ln p(\theta)$, using a numerical optimization routine. Denote the posterior mode by θ^*.

2. Compute the inverse of the negative Hessian at the posterior mode, Σ^*.

3. Initialize the Markov chain by drawing θ^0 from $N(\theta^*, c^2\Sigma^*)$, where c serves as a scaling parameter.

4. For i= 1,...,n_{sim}, draw θ^{prop}, from the proposal distribution $N(\theta^{i-1}, c^2\Sigma^*)$. The jump from θ^{i-1} is accepted (i.e. $\theta^i = \theta^{prop}$) with probability $\min(1, r(\theta^{i-1}, \theta^{prop}|Y)$, where $r(\theta^{i-1}, \theta^{prop}|Y) = \frac{p(Y|\theta^{prop})p(\theta^{prop})}{p(Y|\theta^{i-1})p(\theta^{i-1})}$, and rejected (i.e. $\theta^i = \theta^{i-1}$) otherwise.

The algorithm delivers a sequence of draws $(\theta_i)_{i=1}^{n_{sim}}$ from the density $p(\theta|Y)$. Based on these draws, impulse response functions, density forecasts, or welfare effects for different policy changes can be obtained. Moreover, consistent point estimates can be directly computed for the mean, mode and different quantiles.

Appendix B5. Prediction

Here we reproduce the algorithm that generates draws from the predictive distribution of a DSGE model (Del Negro and Schorfheide [2013]). Denote by T the end of the estimation sample and by h the forecast horizon.

For each posterior draw $\theta^{(i)}$, i= 1,...,n_{sim}, from the posterior distribution:

1. Generate a sequence of innovations $\epsilon^{(i)}_{T+1:T+h}$

2. Starting from s_T^i, iterate forward the transition equation 2.7, with θ replaced by the draw $\theta^{(i)}$ and construct

$$s_t^{(i)} = \Phi_1(\theta^{(i)})s_{t-1}^{(i)} + \Phi_2(\theta^{(i)})\epsilon_t^{(i)}, \quad t = T+1, ..., T+h$$

3. Using the measurement equation given in 2.8, compute n_{sim} trajectories from the predictive distribution.

$$y_t^{(i)} = \Psi_1(\theta^{(i)}) + \Psi_2(\theta^{(i)})s_t^{(i)} \quad t = T+1, ..., T+h$$

The posterior mean forecast $\hat{y}_t$ is obtained by averaging the $y_t^{(i)}$'s.

Appendix C

Appendix to Chapter Three

Appendix C1

Proof of Theorem 1:

Taking a Taylor expansion of the first term of S_P around the probability limit θ_{1R}, we have that:

$$\frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(\hat{u}_{t+h}) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(u_{t+h}) + \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} \nabla_{\theta_1} g(u_{t+h})(\hat{\theta}_{1,R,t} - \theta_{1R}) + o_p(1) \tag{C.1}$$

Denote

$$Z_{1R} = -R^{1/2} J_{1R}^{-1} B^{-1}(\beta_{1R}) DQ_R(\theta_1) \tag{C.2}$$

and

$$\Delta_{1R} = R^{1/2} B(\beta_{1R})(\hat{\theta}_{1R} - \theta_{1R}). \tag{C.3}$$

The matrix $B(\beta_1)$ in (C.2) and (C.3) is defined as $B(\beta_1) = \begin{bmatrix} I_{d_{\psi_1}} & 0_{d_{\psi_1} \times d_{\pi_1}} \\ 0_{d_{\psi_1} \times d_{\pi_1}} & \iota(\beta_1) I_{d_{\pi_1}} \end{bmatrix} \in \mathbb{R}^{d_{\theta_1} \times d_{\theta_1}}$, with $\iota(\beta_1) = \beta_1$ if β_1 is a scalar and $\iota(\beta_1) = \|\beta_1\|$ if β_1 is a vector, and is used to normalize $D^2 Q_R(\theta_{1R})$, so that it is nonsingular asymptotically, when $\beta_{1R} \to 0$ (which is when π_1 is semi-strongly identified).

From the proof of Theorem 3.2 in *AC*2012, that provides the asymptotic distribution of the in-sample estimator of θ, in the semi-strong identification case, we have that:[1]

$$\Delta_{1R} = Z_{1R} + o_p(1) \tag{C.4}$$

Combining (C.2), (C.3) and (C.4) yields:

$$(\hat{\theta}_{1R} - \theta_{1R}) = -J_{1R}^{-1} B^{-2}(\beta_{1R}) D Q_R(\theta_1) + o_p(1) \tag{C.5}$$

Now, in a rolling estimation setup,[2] we can write that for each rolling window $R \leq t \leq T$:

$$(\hat{\theta}_{1,R,t} - \theta_{1R}) = -J_{1R}^{-1} B^{-2}(\beta_{1R}) D^{(t)} Q_R(\theta_1) + o_p(1) \tag{C.6}$$

[1] In the equations below, the index R is the size of the rolling sample.

[2] Note that, in a rolling estimation scheme, for each rolling window indexed by t, $\hat{\theta}_{1,R,t}$ is computed based on the score, $D^{(t)} Q_R(\theta_1)$, indexed by t.

Substituting (C.6) in (C.1) and using Assumption 4, we have:

$$\begin{aligned}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(\hat{u}_{t+h}) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(u_{t+h}) \\ + D_{\theta_1}(-J_1^{-1}(\gamma_{0,1}))\frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} B^{-2}(\beta_{1R})D^{(t)}Q_R(\theta_1) \\ + O_p(\frac{1}{\sqrt{P}}) + o_p(1)\end{aligned} \tag{C.7}$$

as $\frac{1}{P}\sum_{t=R+1}^{T-h} \nabla_{\theta_1} g(u_{t+h}) - D_{\theta_1} = O_p(\frac{1}{\sqrt{P}})$ with $D_{\theta_1} = E(\nabla_{\theta_1} g(u_{t+h}))$.

In a rolling estimation case, the score $D^{(t)}Q_R(\theta_1)$ is of the form $\frac{1}{R}\sum_{j=t-R+1}^{t} m_{1,j}$, where $m_{1,j} \in \mathbb{R}^{d_{\theta_1}}$ is a function of θ_1 and the observables. Thus, we can write:

$$\begin{aligned}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(\hat{u}_{t+h}) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(u_{t+h}) \\ + D_{\theta_1}(-J_1^{-1}(\gamma_{0,1}))B^{-2}(\beta_{1R})\frac{1}{\sqrt{P}R}\sum_{t=R}^{T-h}\sum_{j=t-R+1}^{t} m_{1,j} \\ + O_p(\frac{1}{\sqrt{P}}) + o_p(1)\end{aligned} \tag{C.8}$$

Considering now the second term of the statistic S_P. Analogous to the derivations above we can write that:

$$\begin{aligned}\frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(\hat{\epsilon}_{t+h}) = \frac{1}{\sqrt{P}}\sum_{t=R}^{T-h} g(\epsilon_{t+h}) \\ + D_{\theta_2}(-J_2^{-1}(\gamma_{0,2}))B^{-2}(\beta_{2R})\frac{1}{\sqrt{P}R}\sum_{t=R}^{T-h}\sum_{j=t-R+1}^{t} m_{2,j} \\ + O_p(\frac{1}{\sqrt{P}}) + o_p(1)\end{aligned} \tag{C.9}$$

Equation (C.8) and (C.9) give the expression of S_P,

$$\begin{aligned} S_P = & \frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} (g(u_{t+h}) - g(\epsilon_{t+h})) \\ & + D_{\theta_1}(-J_1^{-1}(\gamma_{0,1}))B^{-2}(\beta_{1R}) \frac{1}{\sqrt{P}R} \sum_{t=R}^{T-h} \sum_{j=t-R+1}^{t} m_{1,j} \\ & - D_{\theta_2}(-J_2^{-1}(\gamma_{0,2}))B^{-2}(\beta_{2R}) \frac{1}{\sqrt{P}R} \sum_{t=R}^{T-h} \sum_{j=t-R+1}^{t} m_{2,j} \\ & + O_p(\frac{1}{\sqrt{P}}) + o_p(1) \end{aligned} \tag{C.10}$$

Now, as all estimators are consistent when π_1 and π_2 are semi-strongly identified, from Theorem 4.1 in West [1996], we have that under the null of $H_0 : E(g(u_{t+h}) - g(\epsilon_{t+h})) = 0$,

$$\frac{1}{\sqrt{P}} \sum_{t=R}^{T-h} (g(\hat{u}_{t+h}) - g(\hat{\epsilon}_{t+h})) \xrightarrow{d} N(0, \Omega)$$

where :

$$\begin{aligned} \Omega = & V_\epsilon + \lambda_1 D'_{\theta_1} V_{W_1} D_{\theta_2} + \lambda_1 D'_{\theta_1} V_{W_2} D_{\theta_2} + 2\lambda_2 D_{\theta_1} C_{\epsilon W_1} - 2\lambda_2 D_{\theta_2} C_{\epsilon W_2} \\ & - 2\lambda_1 D'_{\theta_1} C_{W_1 W_2} D_{\theta_2} \end{aligned}$$

and

$$\begin{aligned} V_\epsilon = & \sum_{j=-\infty}^{+\infty} E[(g(u_{t+h}) - g(\epsilon_{t+h})) - E(g(u_{t+h}) - g(\epsilon_{t+h})) \\ & \times (g(u_{t+h+j}) - g(\epsilon_{t+h+j})) - E(g(u_{t+h}) - g(\epsilon_{t+h}))'] \end{aligned}$$

$$V_{W_1} = -J_1^{-1}(\gamma_{0,1})B^{-2}(\beta_{1R})\Big(\sum_{j=-\infty}^{+\infty} E(m_{1,t}m'_{1,t+j})\Big)B^{-2}(\beta_{1R})(-J_1^{-1}(\gamma_{0,1}))$$

$$V_{W_2} = -J_2^{-1}(\gamma_{0,2})B^{-2}(\beta_{2R})\Big(\sum_{j=-\infty}^{+\infty} E(m_{2,t}m'_{2,t+j})\Big)B^{-2}(\beta_{2R})(-J_2^{-1}(\gamma_{0,2}))$$

$$C_{\epsilon W_1} = \Big(\sum_{j=-\infty}^{+\infty} E[(g(u_{t+h}) - g(\epsilon_{t+h}))m'_{1,t+j}]\Big)B^{-2}(\beta_{1R})(-J_1^{-1}(\gamma_{0,1}))$$

$$C_{\epsilon W_2} = \Big(\sum_{j=-\infty}^{+\infty} E[(g(u_{t+h}) - g(\epsilon_{t+h}))m'_{2,t+j}]\Big)B^{-2}(\beta_{2R})(-J_2^{-1}(\gamma_{0,2}))$$

$$C_{W_1W_2} = -J_1^{-1}(\gamma_{0,1})B^{-2}(\beta_{1R})\Big(\sum_{j=-\infty}^{+\infty} E(m_{1,t}m'_{2,t+j})\Big)B^{-2}(\beta_{2R})(-J_2^{-1}(\gamma_{0,2}))$$

□

Appendix C2: Additional Numerical Results

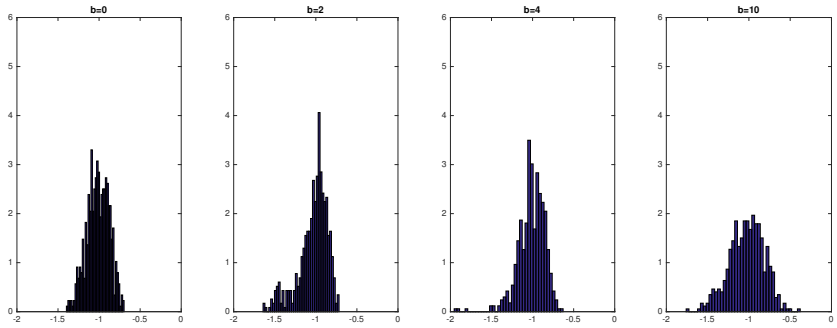

Figure C.1: Finite sample densities of the estimator of ζ_1 when $\pi_0 = -3$.

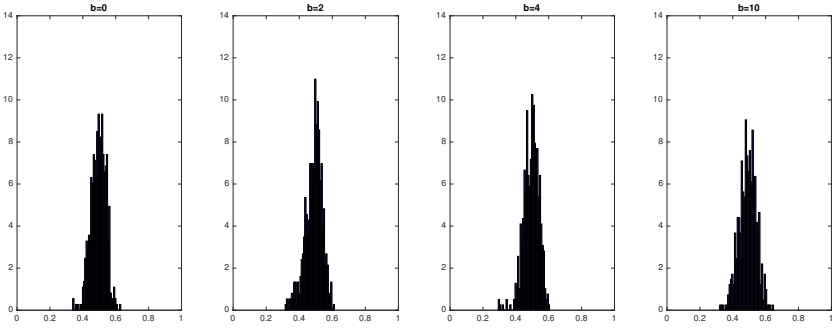

Figure C.2: Finite sample densities of the estimator of ζ_2 when $\pi_0 = -3$.

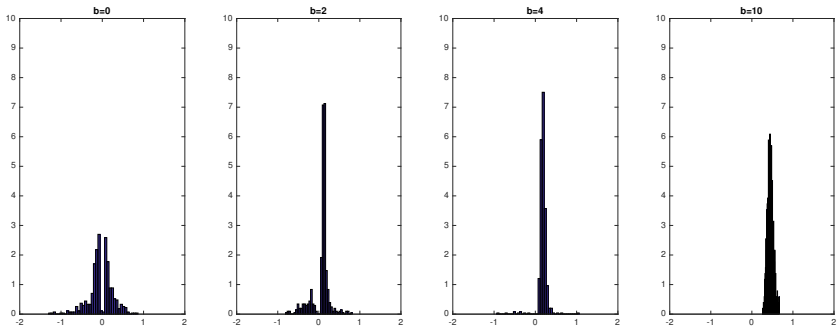

Figure C.3: Finite sample densities of the estimator of β when $\pi_0 = -3$.

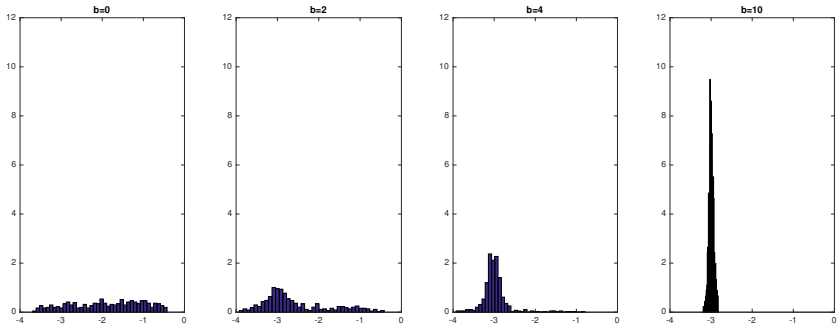

Figure C.4: Finite sample densities of the estimator of π when $\pi_0 = -3$.

Bibliography

Malin Adolfson, Michael Andersson, Jesper Lindé, Mattias Villani, and Anders Vredin. Modern forecasting models in action: Improving macroeconomic analyses at central banks. *International Journal of Central Banking*, 3(4) (111-144), 2007a.

Malin Adolfson, Jesper Lindé, and Mattias Villani. Forecasting performance of an open economy dsge model. *Econometric Reviews*, 26(2-4):289–328, 2007b.

Gianni Amisano and John Geweke. Prediction using several macroeconomic models. *Working Paper, ECB*, 2013.

Gianni Amisano and Raffaella Giacomini. Comparing density forecasts via weighted likelihood ratio tests. *Journal of Business and Economic Statistics*, 25(2):177–190, 2007.

Sungbae An and Frank Schorfheide. Bayesian analysis of dsge models. *Econometric reviews*, 26(2-4):113–172, 2007.

Donald WK Andrews. An empirical process central limit theorem for dependent non-identically distributed random variables. *Journal of Multivariate Analysis*, 38(2):187–203, 1991.

Donald WK Andrews. Estimation when a parameter is on a boundary. *Econometrica*, 67(6):1341–1383, 1999.

Donald WK Andrews. Testing when a parameter is on the boundary of the maintained hypothesis. *Econometrica*, 69(3):683–734, 2001.

Donald WK Andrews and Xu Cheng. Estimation and inference with weak, semi-strong, and strong identification. *Econometrica*, 80(5):2153–2211, 2012.

Donald WK Andrews and Xu Cheng. Maximum likelihood estimation and uniform inference with sporadic identification failure. *Journal of Econometrics*, 173(1):36–56, 2013.

Donald WK Andrews and Xu Cheng. Gmm estimation and uniform subvector inference with possible identification failure. *Econometric Theory*, 30(02): 287–333, 2014.

Donald WK Andrews and Gustavo Soares. Inference for parameters defined by moment inequalities using generalized moment selection. *Econometrica*, 78(1):119–157, 2010.

Donald WK Andrews, Xu Cheng, and Patrik Guggenberger. Generic results for establishing the asymptotic size of confidence sets and tests. 2011.

Isaiah Andrews and Anna Mikusheva. Maximum likelihood inference in weakly identified dsge models. *Quantitative Economics, forthcoming*, 2014.

Isaiah Andrews and Anna Mikusheva. Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics*, 6(1):123–152, 2015.

Isaiah Andrews and Anna Mikusheva. A geometric approach to weakly identified econometric models. *Econometrica, forthcoming*, 2016.

Craig F Ansley and Paul Newbold. Finite sample properties of estimators for autoregressive moving average models. *Journal of Econometrics*, 13(2): 159–183, 1980.

Waldyr Dutra Areosa, Michael McAleer, and Marcelo C Medeiros. Moment-based estimation of smooth transition regression models with endogenous variables. *Journal of Econometrics*, 165(1):100–111, 2011.

Andreas Beyer and Roger EA Farmer. On the indeterminacy of new-keynesian economics. 2004.

Herman J Bierens. Consistent model specification tests. *Journal of Econometrics*, 20(1):105–134, 1982.

Herman J Bierens. A consistent conditional moment test of functional form. *Econometrica*, pages 1443–1458, 1990.

Lev M Bregman. The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming. *USSR computational mathematics and mathematical physics*, 7(3):200–217, 1967.

Fabio Canova and Luca Sala. Back to square one: Identification issues in dsge models. *Journal of Monetary Economics*, 56(4):431–449, 2009.

Fabio Canova, Filippo Ferroni, and Christian Matthes. Approximating time varying structural models with time invariant structures. 2015.

Carlos Capistrán. Bias in federal reserve inflation forecasts: Is the federal reserve irrational or just cautious? *Journal of Monetary Economics*, 55(8): 1415–1427, 2008.

Kung-Sig Chan and Ruey S Tsay. Limiting properties of the least squares estimator of a continuous threshold autoregressive model. *Biometrika*, 85 (2):413–426, 1998.

Saraswata Chaudhuri and Eric Zivot. A new method of projection-based inference in gmm with weakly identified nuisance parameters. *Journal of Econometrics*, 164(2):239–251, 2011.

Xu Cheng. Robust inference in nonlinear models with mixed identification strength. *Journal of Econometrics*, 189(1):207–228, 2015.

Lawrence J Christiano. Solving dynamic equilibrium models by a method of undetermined coefficients. *Computational Economics*, 20(1-2):21–55, 2002.

Kai Philipp Christoffel, Günter Coenen, and Anders Warne. Forecasting with dsge models,. *ECB Working Paper*, 2010.

Peter F Christoffersen and Francis X Diebold. Optimal prediction under asymmetric loss. *Econometric theory*, 13(06):808–817, 1997.

Todd E Clark and Michael W McCracken. Tests of equal forecast accuracy and encompassing for nested models. *Journal of econometrics*, 105(1):85–110, 2001.

Todd E Clark and Michael W McCracken. Evaluating direct multistep forecasts. *Econometric Reviews*, 24(4):369–404, 2005.

Todd E Clark and Michael W McCracken. Tests of equal predictive ability with real-time data. *Journal of Business and Economic Statistics*, 27(4): 441–454, 2009.

Todd E Clark and Michael W McCracken. *Advances in Forecast Evaluation in G. Elliott and A. Timmerman, Handbook of Economic Forecasting*, volume 2. Elsevier, Amsterdam, 2013.

John H Cochrane. Determinacy and identification with taylor rules. *Journal of Political Economy*, (119):565–615, 2011.

Valentina Corradi and Walter Distaso. Multiple forecast model evaluation. *The Oxford Handbook of Economic Forecasting, Oxford University Press, USA*, pages 391–414, 2011.

Valentina Corradi and Norman R Swanson. A consistent test for nonlinear out of sample predictive accuracy. *Journal of Econometrics*, 110(2):353–381, 2002.

Valentina Corradi and Norman R Swanson. Predictive density and conditional confidence interval accuracy tests. *Journal of Econometrics*, 135(1):187–228, 2006.

Valentina Corradi and Norman R Swanson. *Predictive density evaluation in G. Elliott and A. Timmerman, Handbook of Economic Forecasting*, volume 1. Elsevier, Amsterdam, 2006b.

Valentina Corradi and Norman R Swanson. A survey of recent advances in forecast accuracy comparison testing, with an extension to stochastic dom-

inance. In *Recent Advances and Future Directions in Causality, Prediction, and Specification Analysis*, pages 121–143. Springer, 2013.

Valentina Corradi, Andres Fernandez, and Norman R Swanson. Information in the revision process of real-time datasets. *Journal of Business & Economic Statistics*, 27(4):455–467, 2009.

Robert M De Jong. The bierens test under data dependence. *Journal of Econometrics*, 72(1):1–32, 1996.

Marco Del Negro and Frank Schorfheide. Priors from general equilibrium models for vars*. *International Economic Review*, 45(2):643–673, 2004.

Marco Del Negro and Frank Schorfheide. Dsge model-based forecasting. *Handbook of Economic Forecasting, 2a, 57-140, Elsevier*, 2013.

Francis X Diebold. Comparing predictive accuracy, twenty years later: A personal perspective on the use and abuse of diebold–mariano tests. *Journal of Business and Economic Statistics*, 33(1):1–1, 2015.

Jean-Marie Dufour and Mohamed Taamouti. Projection-based statistical inference in linear structural models with possibly weak instruments. *Econometrica*, 73(4):1351–1365, 2005.

Jean-Marie Dufour and Mohamed Taamouti. Further results on projection-based inference in iv regressions with weak, collinear or missing instruments. *Journal of Econometrics*, 139(1):133–153, 2007.

Jean-Marie Dufour, Lynda Khalaf, and Maral Kichian. Identification-robust analysis of dsge and structural macroeconomic models. *Journal of Monetary Economics*, 60(3):340–350, 2013.

Rochelle M Edge, Michael T Kiley, and Jean-Philippe Laforte. A comparison of forecast performance between federal reserve staff forecasts, simple reduced-form models, and a dsge model. *Journal of Applied Econometrics*, 25(4): 720–754, 2010.

Graham Elliott and Ulrich K Müller. Confidence sets for the date of a single break in linear time series regressions. *Journal of Econometrics*, 141(2): 1196–1218, 2007.

Graham Elliott and Ulrich K Müller. Pre and post break parameter inference. *Journal of Econometrics*, 180(2):141–157, 2014.

Graham Elliott, Allan Timmermann, and Ivana Komunjer. Estimation and testing of forecast rationality under flexible loss. *The Review of Economic Studies*, 72(4):1107–1125, 2005.

Graham Elliott, Ivana Komunjer, and Allan Timmermann. Biases in macroeconomic forecasts: irrationality or asymmetric loss? *Journal of the European Economic Association*, 6(1):122–157, 2008.

Jesús Fernández-Villaverde and Juan F Rubio-Ramírez. Estimating macroeconomic models: A likelihood approach. *The Review of Economic Studies*, 74(4):1059–1087, 2007.

Laurent Ferrara, Massimiliano Marcellino, and Matteo Mogliani. Macroeconomic forecasting during the great recession: The return of non-linearity? *International Journal of Forecasting*, 31(3):664–679, 2015.

Franklin M Fisher. *The Identification Problem in Econometrics*. McGraw-Hill, 1966.

Jack Fosten. Forecast evaluation with factor-augmented models. *University of East Anglia, Working Paper*, 2016.

Martin Fukač, Daniel F Waggoner, and Tao Zha. Local and global identification of dsge models: A simultaneous-equation approach. *Working Paper*, 2007.

Eric Ghysels, Arthur Sinko, and Rossen Valkanov. Midas regressions: Further results and new directions. *Econometric Reviews*, 26(1):53–90, 2007.

Raffaella Giacomini. The relationship between dsge and var models. *Advances in Econometrics*, 31, 2013.

Raffaella Giacomini. Economic theory and forecasting: lessons from the literature. *The Econometrics Journal*, 2014.

Raffaella Giacomini and Ivana Komunjer. Evaluation and combination of conditional quantile forecasts. *Journal of Business and Economic Statistics*, 23 (4):416–431, 2005.

Raffaella Giacomini and Giuseppe Ragusa. Theory-coherent forecasting. *Journal of Econometrics*, 182(1):145–155, 2014.

Raffaella Giacomini and Barbara Rossi. Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, 25(4):595–620, 2010.

Raffaella Giacomini and Barbara Rossi. Model comparisons in unstable environments. Technical report, cemmap working paper, 2012.

Raffaella Giacomini and Halbert White. Tests of conditional predictive ability. *Econometrica*, 74(6):1545–1578, 2006.

Tilmann Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106(494):746–762, 2011.

Silvia Goncalves, Michael W McCracken, and Benoit Perron. Tests of equal accuracy for nested models with estimated factors. *FRB St Louis Paper No. FEDLWP2015-025*, 2015.

Pablo Guerron-Quintana, Atsushi Inoue, and Lutz Kilian. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. *Quantitative Economics*, 4(2):197–229, 2013.

Patrik Guggenberger, Frank Kleibergen, Sophocles Mavroeidis, and Linchun Chen. On the asymptotic sizes of subset anderson–rubin and lagrange multiplier tests in linear instrumental variables regression. *Econometrica*, 80(6): 2649–2666, 2012.

Refet S Gurkaynak, Burçin Kisacikoglu, and Barbara Rossi. Do dsge models forecast more accurately out-of-sample than var models? 2013.

James D Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384, 1989.

Bruce E Hansen. Sample splitting and threshold estimation. *Econometrica*, 68(3):575–603, 2000.

Peter Reinhard Hansen. A test for superior predictive ability. *Journal of Business and Economic Statistics*, 2005.

Atsushi Inoue and Barbara Rossi. Testing for weak identification in possibly nonlinear models. *Journal of Econometrics*, 161(2):246–261, 2011.

Nikolay Iskrev. Local identification in dsge models. *Journal of Monetary Economics*, 57(2):189–202, 2010.

Robert G King and Mark W Watson. System reduction and solution algorithms for singular linear difference systems under rational expectations. *Computational Economics*, 20(1-2):57–86, 2002.

Frank Kleibergen and Sophocles Mavroeidis. Weak instrument robust tests in gmm and the new keynesian phillips curve. *Journal of Business and Economic Statistics*, 27(3):293–311, 2009.

Paul Klein. Using the generalized schur form to solve a multivariate linear rational expectations model. *Journal of Economic Dynamics and Control*, 24(10):1405–1423, 2000.

Andrzej Kociecki and Marcin Kolasa. Global identification of linearized dsge models. *Working Paper*, 2014.

Ivana Komunjer and Serena Ng. Dynamic identification of dynamic stochastic general equilibrium models. *Econometrica*, 79(6):1995–2032, 2011.

Ivana Komunjer and Michael T Owyang. Multivariate forecast evaluation and rationality testing. *Review of Economics and Statistics*, 94(4):1066–1080, 2012.

Ivana Komunjer and Yinchu Zhu. Testing the validity of dsge models. *UCSD Working Paper*, 2016.

Gary Koop, M Hashem Pesaran, and Ron P Smith. On identification of bayesian dsge models. *Journal of Business and Economic Statistics*, 31 (3):300–314, 2013.

Tjalling C. Koopmans. *Statistical Inference in Dynamic Economic Models*, volume 10. Wiley New York, 1950.

Jia Li and Andrew J Patton. Asymptotic inference about predictive accuracy using high frequency data. *Economic Research Initiatives at Duke (ERID) Working Paper*, (163), 2013.

Xin Liu and Yongzhao Shao. Asymptotics for likelihood ratio tests under loss of identifiability. *Annals of Statistics*, pages 807–832, 2003.

Ritva Luukkonen, Pentti Saikkonen, and Timo Teräsvirta. Testing linearity against smooth transition autoregressive models. *Biometrika*, 75(3):491–499, 1988.

Leandro M Magnusson and Sophocles Mavroeidis. Identification using stability restrictions. *Econometrica*, 82(5):1799–1851, 2014.

Sophocles Mavroeidis. Identification issues in forward-looking models estimated by gmm, with an application to the phillips curve. *Journal of Money, Credit and Banking*, pages 421–448, 2005.

Michael W McCracken. Asymptotics for out of sample tests of granger causality. *Journal of Econometrics*, 140(2):719–752, 2007.

Jacob A Mincer and Victor Zarnowitz. The evaluation of economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, pages 1–46. NBER, 1969.

Stephen D Morris. Maximum likelihood estimation of globally unidentified dynamic stochastic general equilibrium models using a-priori information. *Working Paper, UCSD*, 2014.

Stephen D Morris and Dan Z. Lee. Local identification of nonlinear dsge models. Technical report, Working Paper, 2014.

Ulrich K Müller. Measuring prior sensitivity and prior informativeness in large bayesian models. *Journal of Monetary Economics*, 59(6):581–597, 2012.

Charles R Nelson and Richard Startz. The zero-information-limit condition and spurious inference in weakly identified models. *Journal of Econometrics*, 138(1):47–62, 2007.

Andrew J Patton. Comparing possibly misspecified forecasts. Technical report, Working paper, 2014.

Andrew J Patton and Allan Timmermann. Properties of optimal forecasts under asymmetric loss and nonlinearity. *Journal of Econometrics*, 140(2): 884–918, 2007a.

Andrew J Patton and Allan Timmermann. Testing forecast optimality under unknown loss. *Journal of the American Statistical Association*, 102(480), 2007b.

Michael K Pitt and Neil Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

Dale J Poirier. Revising beliefs in nonidentified models. *Econometric Theory*, 14(04):483–509, 1998.

Zhongjun Qu. Inference in dynamic stochastic general equilibrium models with possible weak identification. *Quantitative Economics*, 5(2):457–494, 2014.

Zhongjun Qu and Pierre Perron. Estimating and testing structural changes in multivariate regressions. *Econometrica*, 75(2):459–502, 2007.

Zhongjun Qu and Denis Tkachenko. Identification and frequency domain quasi-maximum likelihood estimation of linearized dynamic stochastic general equilibrium models. *Quantitative Economics*, 3(1):95–132, 2012.

Zhongjun Qu and Denis Tkachenko. Local and global parameter identification in dsge models allowing for indeterminacy. *Working Paper*, 2013.

Douglas Rivers and Quang H Vuong. Limited information estimators and exogeneity tests for simultaneous probit models. *Journal of econometrics*, 39(3):347–366, 1988.

Barbara Rossi and Tatevik Sekhposyan. Conditional predictive density evaluation in the presence of instabilities. *Journal of Econometrics*, 177(2): 199–212, 2013.

Thomas J Rothenberg. Identification in parametric models. *Econometrica*, pages 577–591, 1971.

Stephanie Schmitt-Grohé and Martın Uribe. Solving dynamic general equilibrium models using a second-order approximation to the policy function. *Journal of economic dynamics and control*, 28(4):755–775, 2004.

Stephanie Schmitt-Grohé and Martin Uribe. Optimal simple and implementable monetary and fiscal rules. *Journal of Monetary Economics*, 54 (6):1702–1725, 2007.

Frank Schorfheide. Loss function-based evaluation of dsge models. *Journal of Applied Econometrics*, 15(6):645–670, 2000.

Frank Schorfheide, Keith Sill, and Maxym Kryshko. Dsge model-based forecasting of non-modelled variables. *International Journal of Forecasting*, 26 (2):348–373, 2010.

Xiaoxia Shi and Peter CB Phillips. Nonlinear cointegrating regression under weak identification. *Econometric Theory*, 28(03):509–547, 2012.

Christopher A Sims. Solving linear rational expectations models. *Computational economics*, 20(1):1–20, 2002.

Frank Smets and Raf Wouters. An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5):1123–1175, 2003.

Frank Smets and Rafael Wouters. Shocks and frictions in us business cycles: A bayesian dsge approach. *American Economic Review*, 97(586-606), 2007.

Maxwell B Stinchcombe and Halbert White. Consistent specification testing with nuisance parameters present only under the alternative. *Econometric theory*, 14(03):295–325, 1998.

Timo Teräsvirta. Specification, estimation, and evaluation of smooth transition autoregressive models. *Journal of the American Statistical Association*, 89(425):208–218, 1994.

Timo Terasvirta and Heather M Anderson. Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics*, 7(S1):S119–S136, 1992.

H Theil. Economic forecasts and public policy. *Amsterdam, North*, 1958.

Harald Uhlig et al. A toolkit for analysing nonlinear dynamic stochastic models easily. *in: Marimón, R., Scott, A., eds. Computational Methods for the Study of Dynamic Economies. Oxford, UK: Oxford University Press, pp. 30 – 61.*, 1999.

Hal R Varian. A bayesian approach to real estate assessment. *Studies in Bayesian econometrics and statistics in honor of Leonard J. Savage*, pages 195–208, 1975.

Daniel F Waggoner and Tao Zha. Confronting model misspecification in macroeconomics. *Journal of Econometrics*, 171(2):167–184, 2012.

Kenneth D West. Asymptotic inference about predictive ability. *Econometrica: Journal of the Econometric Society*, pages 1067–1084, 1996.

Halbert White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2000.

Arnold Zellner. Bayesian estimation and prediction using asymmetric loss functions. *Journal of the American Statistical Association*, 81(394):446–451, 1986.